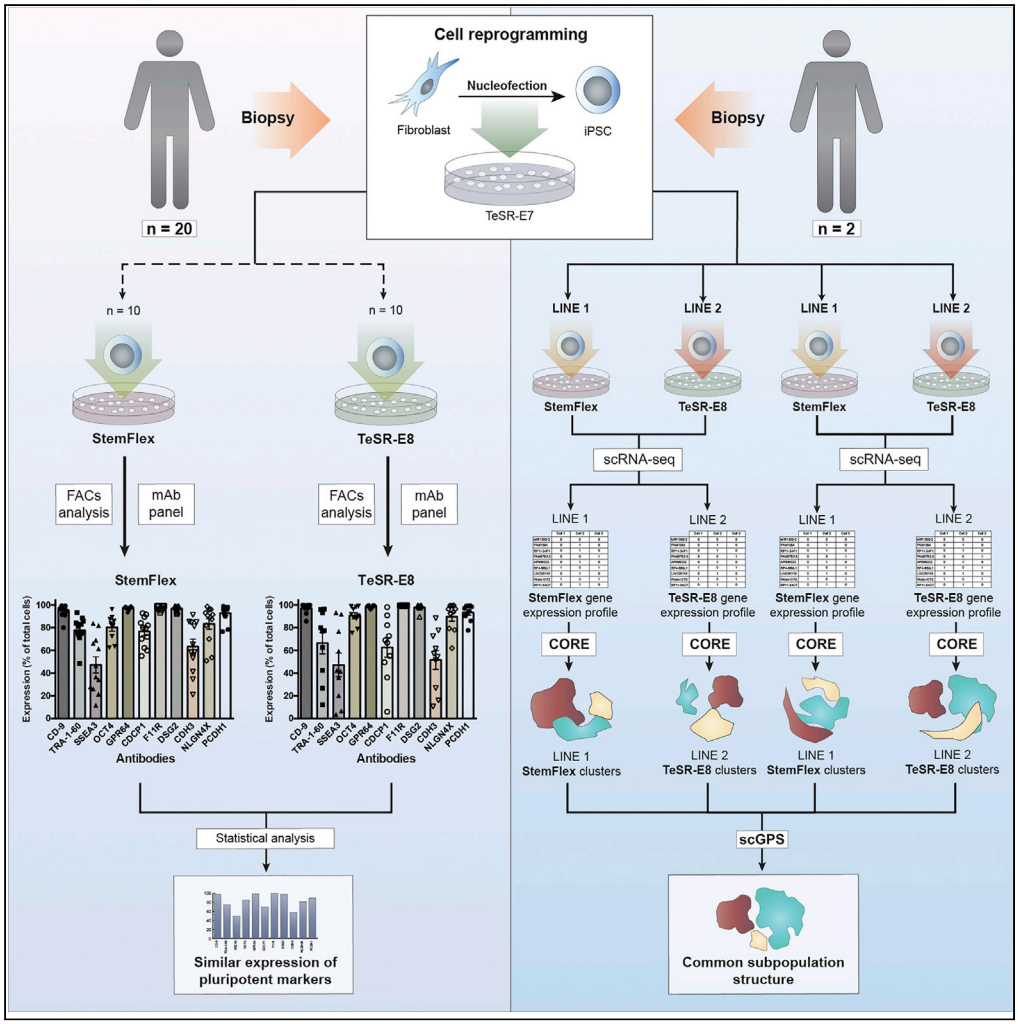
iScience

CellPress

Article

Single-Cell Profiling Identifies Key Pathways Expressed by iPSCs Cultured in Different Commercial Media



Maciej Daniszewski, Quan Nguyen, Hun S. Chy, ..., Alex W. Hewitt, Joseph E. Powell, Alice Pébay

apebay@unimelb.edu.au

HIGHLIGHTS

HiPSCs maintained in StemFlex and E8 were compared by single-cell transcriptomics

Data revealed similar expression of core pluripotency genes

Data revealed a shared three subpopulation structure different in pluripotency states

Daniszewski et al., iScience 7, 30–39
September 28, 2018
https://doi.org/10.1016/j.isci.2018.08.016

iScience

CellPress

Article

Single-Cell Profiling Identifies Key Pathways Expressed by iPSCs Cultured in Different Commercial Media

Maciej Daniszewski,[1,2,11] Quan Nguyen,[3,11] Hun S. Chy,[4,5] Vikrant Singh,[6] Duncan E. Crombie,[1,2] Tejal Kulkarni,[1,2] Helena H. Liang,[1,2] Priyadharshini Sivakumaran,[1,2] Grace E. Lidgerwood,[1,2] Damián Hernández,[1,2] Alison Conquest,[1,2] Louise A. Rooney,[1,2] Sophie Chevalier,[1,2] Stacey B. Andersen,[3] Anne Senabouth,[3] James C. Vickers,[7] David A. Mackey,[8] Jamie E. Craig,[9] Andrew L. Laslett,[4,5] Alex W. Hewitt,[1,2,6,11] Joseph E. Powell,[3,10,11] and Alice Pébay[1,2,11,12,*]

SUMMARY

We assessed the pluripotency of human induced pluripotent stem cells (iPSCs) maintained on an automated platform using StemFlex and TeSR-E8 media. Analysis of transcriptome of single cells revealed similar expression of core pluripotency genes, as well as genes associated with naive and primed states of pluripotency. Analysis of individual cells from four samples consisting of two different iPSC lines each grown in the two culture media revealed a shared subpopulation structure with three main subpopulations different in pluripotency states. By implementing a machine learning approach, we estimated that most cells within each subpopulation are very similar between all four samples. The single-cell RNA sequencing analysis of iPSC lines grown in both media reports the molecular signature in StemFlex medium and how it compares to that observed in the TeSR-E8 medium.

INTRODUCTION

Somatic cells can be reprogrammed into induced pluripotent stem cells (iPSCs) using a combination of transcription factors involved in the maintenance of pluripotency (Park et al., 2007; Takahashi and Yamanaka, 2006; Takahashi et al., 2007; Yu et al., 2007). Like human embryonic stem cells (hESCs), human iPSCs provide a powerful means by which to investigate the pathogenesis of disease because these cells can be differentiated into relevant cell types of interest for disease modeling, drug screening, and delving into the fundamental aspects of disease pathology, as well as for regenerative medicine. Numerous protocols have been described for the maintenance of human pluripotent stem cells, using different matrices and culture media (International Stem Cell Initiative Consortium et al., 2010). The original method of maintenance of hESCs was based on a feeder layer of embryonic fibroblasts and a serum-based medium (Reubinoff et al., 2000; Thomson, 1998). It evolved to more defined conditions using serum-free media with key signaling molecules (Furue et al., 2008; Pébay et al., 2005; Vallier et al., 2005; Xu et al., 2005) and/or feeder-free conditions (including the use of Matrigel, collagen, vitronectin, or laminin) (Klimanskaya et al., 2005; Ludwig et al., 2006). Various commercialized serum-free media have now become available with an ongoing evolution to fully defined media, from KnockOut Serum Replacement medium supplemented with basic fibroblast growth factor to the TeSR family media (mTeSR1, TeSR2, and TeSR-E8) (Chen et al., 2011; Ludwig et al., 2006; Sun et al., 2009), and the newly released StemFlex, a medium with proprietary compounds, and hence of unknown formulation and with robustness efficiency remaining to be characterized.

Although there have been clear advances in the maintenance, standardization, and upscaling of pluripotent stem cell culture, some limitations remain and need to be addressed to realize the translational potential of iPSCs (Kim and Kino-oka, 2018). Indeed, human variability is a main source of differences observed between cell lines generated and maintained in various laboratories (Allegrucci and Young, 2006; Allegrucci et al., 2007), with genetic background and methodologies also having a significant contribution (Kilpinen et al., 2016). The automation of pluripotent stem cell culture provides an efficient way to improve these aspects, increasing throughput and standardizing many aspects of cell culture, including in iPSC generation, maintenance, passaging, and differentiation into progeny cells of interest (Crombie et al., 2017; Konagaya et al., 2015; Paull et al., 2015). These automated steps allow minimal variation in each procedure,

[1]Centre for Eye Research Australia, Royal Victorian Eye and Ear Hospital, University of Melbourne, 32 Gisborne Street, East Melbourne, VIC 3002, Australia

[2]Ophthalmology, Department of Surgery, the University of Melbourne, Melbourne, VIC 3002, Australia

[3]Institute for Molecular Bioscience, University of Queensland, Brisbane, QLD 4072, Australia

[4]Commonwealth Scientific and Industrial Research Organisation (CSIRO) Manufacturing, Clayton, VIC 3168, Australia

[5]Australian Regenerative Medicine Institute, Monash University, Clayton, VIC 3168, Australia

[6]School of Medicine, Menzies Institute for Medical Research, University of Tasmania, Hobart, TAS 7000, Australia

[7]Wicking Dementia Research and Education Centre, University of Tasmania, Hobart, TAS 7000, Australia

[8]Centre for Ophthalmology and Visual Science, Lions Eye Institute, University of Western Australia, Perth, WA 6009, Australia

[9]Flinders University, Adelaide, SA 5042, Australia

[10]Garvan Institute of Medical Research, Darlinghurst, Sydney, NSW 2010, Australia

[11]These authors contributed equally

[12]Lead Contact

Continued

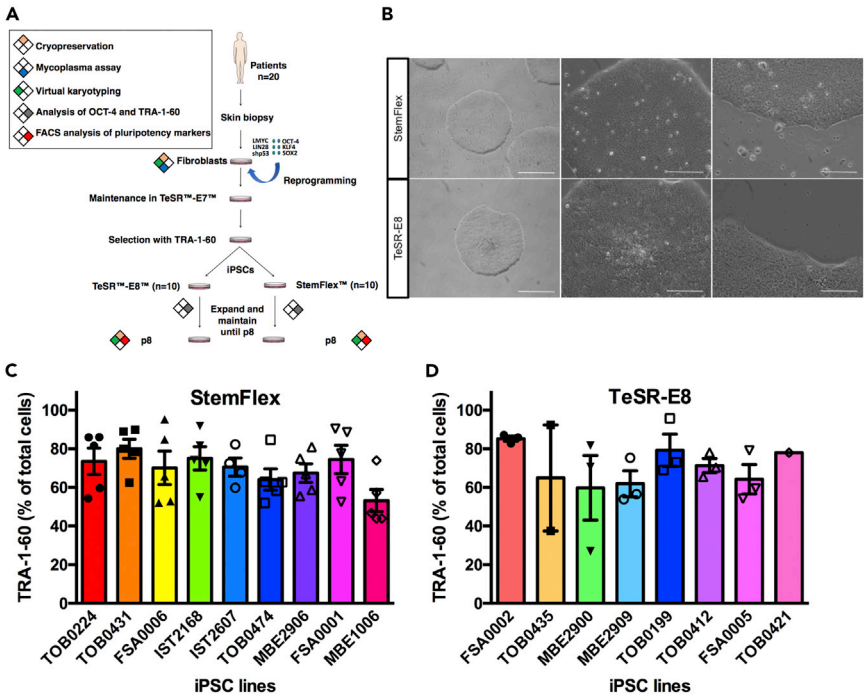


Figure 1. Quality Control of iPSCs Cultivated in StemFlex and TeSR-E8

(A) Schematic representation of the experimental plan.

(B) Representative morphology of iPSCs maintained in StemFlex or TeSR-E8 post-passaging. Images of human iPSCs, at day 8 post-passaging, at different magnifications (×4, ×10, and ×20). Cells were cultivated using an automated platform, on vitronectin, in StemFlex (IST2607, p3 1:6 ratio) or TeSR-E8 (TOB0435, p2 1:3 ratio). Images are representative of all cell lines. Scale bars: 100 μM.

(C and D) Quantification of TRA-1-60-positive cells before passaging of human iPSCs maintained in StemFlex (C) or TeSR-E8 (D). Scatterplot with bar of the TRA-1-60-positive (+ve) cells just before passaging, in independent iPSC lines grown in StemFlex (TOB0224 p3,5–8; TOB0431 p3,5–8; FSA0006 p3,5–8; IST2168 p3,5–8; IST2607 p3,5,6,8; TOB0474 p3,5–8; MBE2906 p3,5–8; FSA0001 p2,4–6,8; MBE1006 p2,4–6,8) or TeSR-E8 (FSA0002 p2,5,8; TOB0435 p5,8; MBE2900 p2,5,8; MBE2909 p4,5,8; TOB0199 p2,4,8; TOB0412 p3,4,8 FSA0005 p4,5,8 TOB0421 p8). The abnormal lines TOB0218 (StemFlex), MBE2899 (TeSR-E8), and MBE2901 (TeSR-E8) were not included in this analysis. Each column represents the mean ± SEM of successive passages of individual lines. Two-tailed t test indicates no statistical difference in TRA-1-60 expression between the two culture conditions ($p = 0.8546$).

reduce inter-sample variability, and hence increase robustness of cell culture procedures (Daniszewski et al., 2017).

Using an automated platform to ensure robustness of cell culture, we compared the cellular and molecular profiles of human iPSCs maintained in a standard TeSR-E8 medium and in the newly released StemFlex. Quality control was assessed by markers of pluripotency, three germ layer differentiation, and virtual karyotyping. We then compared samples grown in both media by single-cell RNA sequencing (scRNA-seq) to uncover the molecular signature underlying pluripotency, and report on whether similar pathways are modulated in both conditions.

RESULTS

Donor iPSCs were derived as described in Crombie et al. (2017). Following TRA-1-60 selection, 10 lines were maintained in StemFlex medium and 10 lines in TeSR-E8 medium (Figure 1A). We had already adapted automation for the culture of pluripotent stem cells using TeSR-E8 (Crombie et al., 2017), and StemFlex also allowed maintenance on the automated platform, with colonies showing typical pluripotent stem cell morphology (Figures 1B and S1) and similar levels of expression of TRA-1-60 (Figures 1C and 1D). After 8 passages, all iPSC lines for both media expressed the pluripotency markers OCT-4 and TRA-1-60 (Figures 2A–2L). All iPSC lines in both conditions were able to differentiate to the three germ layers, as

*Correspondence:
apebay@unimelb.edu.au

https://doi.org/10.1016/j.isci.2018.08.016

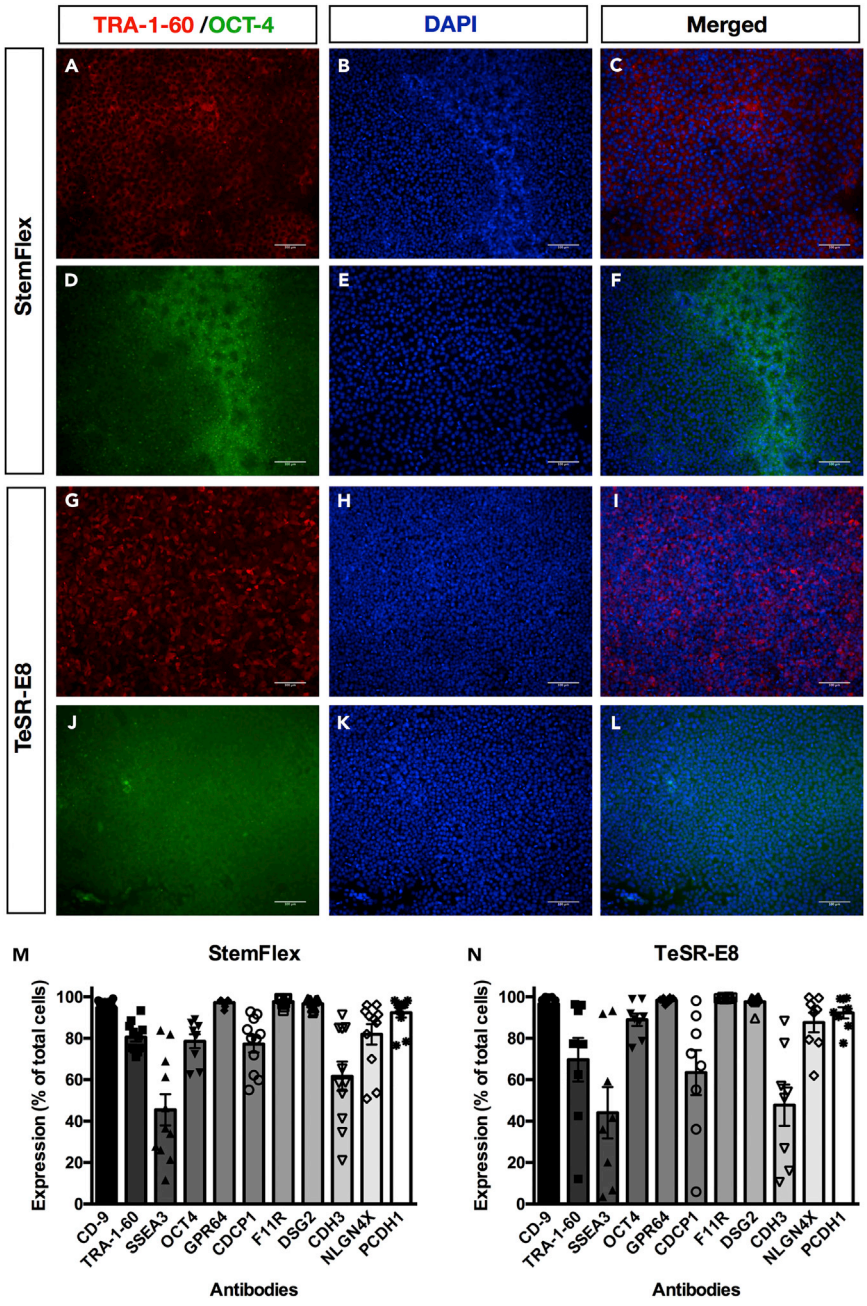


Figure 2. Expression of Pluripotency Markers of Human iPSCs Maintained in StemFlex or TeSR-E8

(A–L) Representative immunostainings for TRA-1-60 and OCT-4. Images of human iPSCs cultivated in StemFlex (TOB0474) or TeSR-E8 (MBE2901) at day 8 post-passaging (p8), for TRA-1-60 (A and G) or OCT-4 (D and J), with DAPI nuclei counterstain (B, E, H, and K) and merged images (C, F, I, and L). Images are representative of all cell lines. Scale bars: 100 μM.

(M and N) Quantification of pluripotency markers. Scatterplot with bar of human iPSCs cultivated in (M) StemFlex (TOB0224; TOB0431; FSA0006; IST2168; IST2607; TOB0474; MBE2906; FSA0001; MBE1006) or (N) TeSR-E8 (TOB0421; FSA0005; TOB0199; FSA0002; MBE2909; TOB0435; MBE2900; TOB0412) at day 8 post-passaging (p16), for CD9, TRA-1-60, SSEA-3, OCT-4, GPR64, CDCP1, F11R, DSG2, CDH3, NLGN4X, and PCDH1. Each column represents the mean ± SEM of 10 individual lines. FSA0006 and TOB0474 were analyzed twice for all markers aside from OCT-4, which was analyzed once. The abnormal lines TOB0218 (StemFlex), MBE2899 (TeSR-E8), and MBE2901 (TeSR-E8) were not included in this analysis. Two-way ANOVA followed by Sidak's multiple comparisons test indicates no statistical difference in each pluripotency marker expression between the two culture conditions.

assessed by embryoid body formation (Figure S2). Virtual karyotyping using SNP analysis revealed post-reprogramming anomalies in two lines grown in TeSR-E8 (MBE2899, MBE2901) and one line maintained in StemFlex (TOB0218) (Figure S3). At passage 16, iPSCs grown in both media were subjected to flow cytometry for a panel of monoclonal antibodies that detect human pluripotent stem cells (O'Brien et al., 2017), increasing the stringency of the characterization of cell pluripotency markers. Cells retained high levels of expression of the panel of pluripotent markers in both media (Figures 2M and 2N). Of note, SSEA-3 was found to be low in all samples; however, this marker is known not to be essential to pluripotency (Brimble et al., 2007).

We interrogated the molecular pathways involved in StemFlex and TeSR-E8 maintenance of pluripotency. We reprogrammed two independent donor fibroblast lines, WAB0450 and WAB0069; selected iPSCs using TRA-1-60 beads; and separated each iPSC line into the two media post-passage 1 (Figure 3A). Following virtual karyotyping at passage 8 to ensure that cells did not acquire chromosomal abnormalities (Figures S4A–S4F), and for confirmation of pluripotency by embryoid body formation (Figure S2), iPSCs were harvested for scRNA-seq to identify potential variation in pluripotency, associated pathways, and potential culture medium signature. Passage 8 was chosen because the number of aberrant cells with aberrant copy number variation (CNVs) is already reduced from early passages (Hussein et al., 2011). At day 5 post-passaging, iPSCs were dissociated into single cells, and live cells were selected with propidium iodide by fluorescence-activated cell sorting (FACS). Live cells were used for library preparation and scRNA-seq, as outlined in Figure 2A. Processing of our initial analysis identified a total of 21,597 cells in four samples. We filtered a total of 635 cells, which met one or more of the following criteria: high/low mapped reads (163 cells), high mitochondrial expression (456 cells), and high ribosomal expression (57 cells), resulting in 20,962 cells remaining for subsequent analysis. We mapped reads to a reference transcriptome (GRCh38p10) containing 32,838 genes, of which we filtered 16,459 because they were detected in less than 0.1% of the total number of cells leaving 16,379 genes for the analysis. The filtered data were normalized, and cells were then divided into individual samples and clustered separately. The distribution of cells based on gene expression profile showed a clear overlap between the four culturing samples, suggesting the overall similar effects of the two media (Figure 3B). For identifying genes and pathways involved in maintaining pluripotency, we performed differential gene expression (DE) between media, assessing both individual lines and pooled lines: WAB0450_E8 was analyzed together with WAB0069_E8 and WAB0450_SF together with WAB0069_SF followed by analysis of individual lines (Table S1). Furthermore, we performed both targeted expression analysis of 53 known pluripotency markers (Figure 3C) and genome-wide expression comparisons for all 16,379 genes (Figure 3D) in the four separate samples or in the pooled media. Gene expression was compared both by expression values (Figures 3C and 3D) and by percent of expressing cells (Tables S1 and S2, Figure S5). Cell cultures retained expression of the panel of pluripotent cell surface markers (Table S2). Notably, 99.96% of all cells in TeSR-E8 and 99.97% of all cells in StemFlex expressed at least one of the three core pluripotency markers, namely, *POU5F1*, *SOX2,* and *NANOG*, suggesting that cells are maintained in a pluripotency state (Table S2, Figure S5). The percentage of cells expressing *NANOG* is at 75.5th percentile of all 16,270 reliably detected genes, higher than 13,349 genes. The distribution of number of detected cells for every gene suggests that *NANOG* was detected in more cells than most other genes (Figure S6). Although the total number of up- and downregulated genes in cells in the two media are similar (Figures 3C and 3D), a number of differentially regulated genes were unique for cells in each medium (Table S1). Pluripotency markers expressed higher in cells maintained in StemFlex include *LEFTY1*, *LEFTY2*, *IFITM1*, *SFRP2*, *REST*, *OTX2*, *TCF3*, *BRIX1*, *KLF4*, *KLF5*, *HESX1*, *CRABP2*, *NR5A2*, *FOXD3*, *DNMR3B*, *UTF1*, and *PTEN* (Figure 3C). These genes are particularly enriched for "Signaling by NODAL" (Reactome pathway analysis, false discovery rate [FDR] $< 4 \times 10^{-3}$). Gene markers expressed more in cells maintained in TeSR-E8 include *TCL1B*, *CD7*, *GDF3*, *LIFR*, *GBX2*, *CXCL5*, *CDH1*, *FGF4*, *GAL*, *SOX2*, *POU5F1*, *DPPA2*, *PODXL*, *IFITM2*, *NANOG*, *ZFP42*, *TCL1A*, *NODAL*, *DPPA5*, *COMMD3*, *SEMA3A*, *POU3F1*, *PRDM14*, and *SALL4* (Figure 3C). These TeSR-E8-upregulated genes are enriched for "*POU5F1*, *SOX2*, *NANOG* repress genes related to differentiation and activate genes related to proliferation" (Reactome pathway analysis, FDR $< 1 \times 10^{-4}$). Notably, both sets of upregulated genes share the main enriched pathway "Transcriptional regulation of pluripotent stem cells" (Reactome pathway analysis, FDR $< 4 \times 10^{-3}$ and FDR $<1.9 \times 10^{-15}$) (Figure 3C, Tables S1 and S2, Figure S7). We observed that the combination of 18 markers associated with the naive pluripotency (Collier et al., 2017; Guo and Smith, 2010; Ng and Surani, 2011; Warrier et al., 2017; Weinberger et al., 2016) were expressed by a low but comparable percentage (with difference lower than 1%) of cells grown in both conditions (Table S2). Moreover, genes driving transition from naive to primed pluripotency, i.e., *FGF4*, *FOXD3,* and *OTX2* (Collier et al., 2017) were also

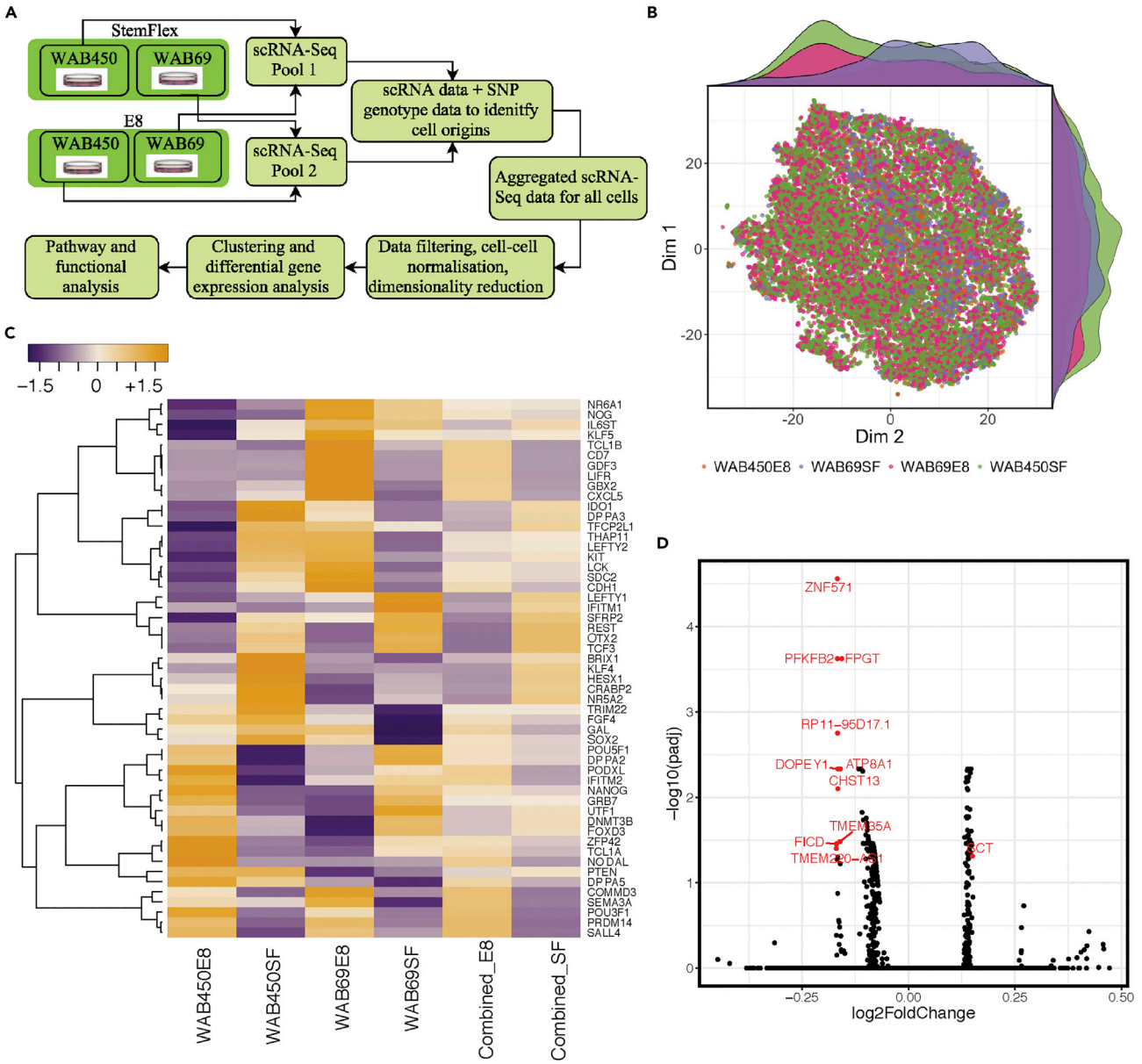


Figure 3. Comparing Effects of the Two Media at Single-Cell Level

(A) Schematic representation of the experimental randomization design. Two cell lines were cultured separately in two media. The four cell cultures were genotyped (HumanCore Beadchip arrays) and were randomly combined into two pools for scRNA-seq experiment (10X Chromium). Cells were assigned to original sample based on sample SNP genotype and single-cell SNPs called from scRNA-seq data. The scRNA-seq data were aggregated for all four samples, filtered, normalized, clustered, and analyzed for differential gene expression and functional pathways.

(B) Two-dimensional distribution of cells based on gene expression profile. Dimensions 1 and 2 are principal coordinates of imputed values from CIDR (Clustering through Imputation and Dimensionality Reduction) (Lin et al., 2017). The overlap between the four culturing samples suggests the overall similar effects of the two media.

(C) Heatmap analysis of known pluripotency markers shown for four separate samples or the combination of two samples. Standardized gene expression is shown from low (purple color) to high (orange color). The numbers of genes expressed highly in StemFlex or in TeSR-E8 are similar, suggesting the two media maintain similar pluripotency states.

(D) Volcano plot shows genome-wide differential expression results between cells in StemFlex versus cells in TeSR-E8 (positive log2Folchange indicates higher expression in StemFlex). The numbers of upregulated genes in each media are similar. Red text indicates genes for which upregulation is statistically significant between the two conditions.

identified in subpopulations from all samples; however, on average the difference between the two media did not exceed 2% (Table S2). We also found high expression of genes involved in "Proliferation and survival" pathways and in "Metabolism" pathways (Table S2).

We then performed high-resolution comparisons at subpopulation and single-cell levels. Through clustering of the 20,962 cells, we consistently identified three distinct subpopulations in each of the WAB0450_SF, WAB0450_E8, WAB0069_E8, and WAB0069_SF samples (Figures S7–S9, Tables S4 and S5). The subpopulations were named 1, 2, and 3 based on the order of the total cell counts, with subpopulation 1 having the highest count. The dendrogram trees (Figures S7 and S8) show the clustering results, from applying our unbiased clustering method described previously (Nguyen et al., 2018). Statistical analysis of DE between the identified subpopulations, followed by pathway enrichment analysis of significantly regulated genes, suggests two major distinct pluripotency states. Subpopulations 1 and 2, which together account for 77%–95% of total cells in a sample (Figures S7–S9), had gene expression signature closer classifying them as more pluripotent cells compared with cells in subpopulation 3, which are more primed to differentiation (5%–23%, Tables S3 and S6). Table S3 shows enriched pathways for upregulated genes when comparing subpopulations 1 and 2 to subpopulation 3. The comparison was performed separately for each of the four samples. This comparison allows the independent assessment of the three subpopulations in each cell line and each medium. In addition, Table S6 shows functional pathway analysis for the common gene markers expressed highest in each of the three subpopulations and found consistently in all four samples. This comparison combines the shared pattern found in all four samples, thereby reinforcing the potential gene markers for each of the three subpopulations. Overall, the enrichment of pluripotency for subpopulations 1 and 2 compared with subpopulation 3 is consistent across cell lines and media (Tables S3 and S6), which is consistent with results from our previous study of a human iPSC population (Nguyen et al., 2018). Specifically, subpopulations 1 and 2 had a higher expression level of pluripotency markers, with significant enrichment (p adjusted < 0.001) of the Reactome pathway "Upregulation of transcriptional regulation of pluripotent stem cells" (>20 genes in the pathway were significantly upregulated, Table S3). For example, consistently in four samples (WAB0450_SF, WAS0450_E8, WAB0069_SF, and WAB0069_E8), the RNA levels of key pluripotency gene markers such as *POU5F1*, *SOX2*, and *NANOG* were higher in cells within subpopulations 1 and 2 than in cells within subpopulation 3 (Table S3). Notably, the percent of cells expressing these markers were higher in subpopulation 3 than in the remaining cells, possibly due to the relatively smaller number of cells present in this subpopulation (5%–23% of the total cells). Compared with subpopulations 1 and 2, the cells in subpopulation 3 were more primed to differentiation, with the significant upregulation (FDR <0.05) of differentiation-promoting pathways, such as "Signaling by NOTCH4" (FDR $<3.6 \times 10^{-7}$), "RUNX1 regulates transcription of genes involved in differentiation of HSCs" (FDR $<1.5 \times 10^{-5}$), "Beta-catenin-independent WNT signaling" (FDR <0.0003), and "TCF dependent signaling in response to WNT" (FDR <0.008, Table S6). The complete list of all gene markers that are expressed highest in each subpopulation are shown in Table S6.

To compare subpopulation structure between two independent samples (e.g., same cell line in two media, or different cell lines in two media), we applied two approaches. First, we performed unbiased clustering of cells in each of the sample and compared the clustering results between separate samples and between individual sample with the combined set of all cells from four samples (Figures S7–S10). The grouping of similar cells into clusters (subpopulations) reflects the structure of the cell population in the original population, because the clustering is data driven and does not require prior knowledge and assumptions on the number of clusters. We consistently found three subpopulations in each of the four samples, suggesting the shared subpopulation structure (Figures S7 and S8). We found that the distribution of cells into cell-cycle stages is similar between the four samples, suggesting that the subpopulation structure is not driven by cell-cycle phases (Figure S9). Moreover, we performed clustering analysis for the merged dataset of all cells from all four samples (2 cell lines × 2 media). We also found three subpopulations, and importantly that the percent of cells allocated to subpopulations 1, 2, and 3 are similar between all four samples, suggesting that the subpopulation structure is independent of cell lines and culturing media (Figures S7 and S8). Second, to quantitatively compare between subpopulations identified separately in each of the four samples, we performed a machine learning model, which estimates pairwise transition scores between subpopulations from one sample to another. For each pair of subpopulations identified through clustering, we run scGPS (single-cell global fate potential of subpopulation), which is a machine learning classification method to compare every cell from one subpopulation to another subpopulation, as described in Nguyen et al. (2017). This method was applied to quantitatively estimate conditional class probability of every single

cell belonging to a subpopulation, thus allowing to estimate the percent of similar cells between two subpopulations. We found that approximately 98% of the cells from subpopulation 3 in one sample were similar to cells in subpopulation 3 from the other sample, much higher than the similarity when compared with cells in subpopulations 1 and 2 (Table S4). Thus, the analysis suggests that all samples share the common subpopulation 3. Similarly, subpopulations 1 of one cell line (WAB0069) or two different cell lines in StemFlex and TeSR-E8 media are 95.0% and 95.8% similar, suggesting subpopulations 1 are the same among these samples (Table S4). Notably, subpopulations 1 and 2 in the cell line WAB00450_E8 are less specific and have the potential to convert into subpopulations 1, 2, and 3. Interestingly, subpopulations 2 in WAB0450_E8 and WAB0069_E8 samples display an intermediate state and have a high (ranging from 82% to 100%) similarity score (i.e., high potential to be the same cell types/cell state) compared with subpopulations 2 or 3 in WB0069_SF (Table S4). Overall, we found a shared subpopulation structure among all samples, in which subpopulations 1 and 3 are more discrete and present in all samples, whereas subpopulation 2 is more transient and has the potential to dynamically transition between subpopulations 2 and 3.

DISCUSSION

We analyzed the ability of the recently released StemFlex and of TeSR-E8, a widely used cell medium, to maintain pluripotency of newly reprogrammed human iPSCs. We used an automated platform to perform this analysis to ensure that variations observed in cell maintenance were independent of human variability. All lines maintained in each medium (10 independent lines per media) demonstrated robust maintenance of pluripotency, with key colony morphology and expression of pluripotency markers, analyzed by immunochemistry and FACS analysis at passage 16. Most lines retained a normal karyotype with the exception of two in TeSR-E8 and of one in StemFlex. Our data thus support the robustness of both media to maintain iPSCs, as cells grown in both conditions display the morphological and molecular characteristics of iPSCs. In-depth molecular analysis was then performed using scRNA-seq to access the molecular network in place in cells grown in both culture conditions. Two independent fibroblast cultures were reprogrammed, and the two generated iPSC lines were then maintained and passaged in both media in parallel. Following virtual karyotyping, cells from different cell lines and media were then combined before single-cell sequencing. This study design was chosen to limit the inter-variability observed between independent iPSC lines. The analysis of gene expression profiles in the four samples—two lines each maintained in the two media—indicate that they share a very similar global gene expression signature, although specific pathways are enriched for cells in each of the two media. Our analysis identified that the core pluripotency factors, i.e., *OCT-4*, *NANOG*, and *SOX2*, were expressed by similar proportions of cells in both StemFlex and TeSR-E8 (Table S2), and >99.9% of all cells expressed at least one of these three factors, which is an indication of potency of both media to maintain cells in a pluripotent state, as levels of expression of those transcription factors are tightly correlated and critical to maintain pluripotency (Rizzino, 2009). At individual gene levels, we observed that on average 18.1% and 19.5% of TeSR-E8- and StemFlex-cultured cells, respectively, expressed *NANOG* (range 9.9%–44.5%; Tables S2 and S5). Its heterogeneous expression was reported previously both in mouse embryonic stem cells (Chambers et al., 2007) and human iPSCs (Yu et al., 2018), and it was suggested that *NANOG* levels in cells exist in a state of dynamic equilibrium (Boyer et al., 2005; Loh et al., 2006). The low levels of *NANOG* found in this study are consistent with previous results from an scRNA-seq iPSC dataset, showing *NANOG* expression level to be approximately 47 times lower than that of *OCT-4* (Nguyen et al., 2018). *OCT-4* is also essential for acquisition of naive pluripotency, and for differentiation (Radzisheuskaya et al., 2013). Furthermore, *OCT-4* and *SOX2* co-regulate their expression (Chew et al., 2005), while they jointly regulate *NANOG* (Rizzino, 2009).

Altogether, our data indicate that StemFlex is an efficient medium for the maintenance of human iPSCs. StemFlex-cultured cells retain characteristic iPSC morphology with round-shaped, tight morphology of colonies and expression of pluripotency markers. Little difference in global gene expression signatures between iPSCs cultured in StemFlex and TeSR-E8 could suggest that their formulations are highly similar, and also highlights the advantage of automation in stem cell research where eliminating human variability improves consistency and robustness of culture. Of note, automated passaging using ReLeSR was more efficient for cells maintained in StemFlex than in TeSR-E8. In our automation procedures, detaching cells from culture plates was faster with StemFlex than with TeSR-E8 and rarely required an additional "offline" step of "taping," required for TeSR-E8. This was selected as an advantage of StemFlex for automation, for decreasing passaging time, reducing human intervention, and improving passaging efficiency. Our data also indicate that iPSCs cultured in both media exist in a metastable heterogeneous state supporting

self-renewal, which is maintained by external cues within the culture media, in concordance with previous reports in mouse (Hanna et al., 2009; Najm et al., 2011) and human pluripotent stem cells (Buecker et al., 2010; Hough et al., 2014). These data also describe the single-cell resolution molecular pathways modulated in StemFlex and how they compare to those modulated in TeSR-E8.

Limitations of Study

There are some limitations to this study. First, conducting the analysis at several time points would likely generate additional information regarding how signaling pathways of stem cells may be affected by prolonged culture in artificial *in vitro* conditions. The single time point used for the transcriptome analysis of single cells restricts the understanding of expression analysis over time. The repeat of such analysis at various passaging times could provide further information on the stability of cells in culture in both media. Furthermore, overall, the scRNA-seq dataset had relatively low number of mapped reads per cells: less than 20,000 unique molecular identifiers per cell (Figure S10). We addressed the limitation by utilizing the large number of cells (20,962 cells) and a stringent filtering step to remove cells with a low number of mapped reads and genes detected in a small percentage of cells. Another limitation is the small number of biological replicates used in the single-cell sequencing experiment, with two cell lines (from two genetic backgrounds) per medium. This could result in low detection power to reveal the effect derived from variation in genetic backgrounds from the scRNA-seq dataset. We complemented this deficit by using more cell lines, with 10 different genetic backgrounds for each medium, in the flow cytometry panel experiments. The cross-comparison between the two experimental approaches showed consistent results.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

DATA AND SOFTWARE AVAILABILITY

We have made available both the raw and normalised data on ArrayExpress under accession E-MTAB-6524.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Transparent Methods, 10 figures, and 6 tables and can be found with this article online at https://doi.org/10.1016/j.isci.2018.08.016.

ACKNOWLEDGMENTS

We would like to thank Erik Willems, Rhonda Newman, and Jennifer Crean from Thermo Fisher Scientific for their support in this work. This work was supported by grants from Stem Cells Australia—the Australian Research Council Special Research Initiative in Stem Cell Science, the Ophthalmic Research Institute of Australia, Retina Australia, the Joan and Peter Clemenger Foundation, the Philip Neal bequest, a National Health and Medical Research Council Practitioner Fellowship (AWH), an Australian Research Council Future Fellowship (A.P., FT140100047), National Health and Medical Research Council Project Grant (J.C.V.), an International Postgraduate Research Scholarship & Research Training Program Scholarship (M.D.), the University of Melbourne, and Operational Infrastructure Support from the Victorian Government.

AUTHOR CONTRIBUTIONS

M.D., Q.N.: concept and design, experimental work, interpretation of data, writing of manuscript, final approval of manuscript; H.S.C.: experimental work, interpretation of data, writing of manuscript, final approval of manuscript; D.E.C., V.S., P.S., G.E.L., T.K., H.H.L., A.C., D.H., L.A.R., S.C., S.B.A., A.S.: experimental work, interpretation of data, final approval of manuscript; J.C.V., D.A.M., J.E.C.: provision of samples, final approval of manuscript; A.L.L.: concept and design; interpretation of data, final approval of manuscript; J.E.P., A.W.H., A.P.: concept and design, financial support, interpretation of data, writing of manuscript, final approval of manuscript.

DECLARATION OF INTERESTS

None.

Received: March 13, 2018
Revised: August 9, 2018
Accepted: August 17, 2018
Published: September 28, 2018

REFERENCES

Allegrucci, C., and Young, L.E. (2006). Differences between human embryonic stem cell lines. Hum. Reprod. Update *13*, 103–120.

Allegrucci, C., Wu, Y.-Z., Thurston, A., Denning, C.N., Priddle, H., Mummery, C.L., Ward-van Oostwaard, D., Andrews, P.W., Stojkovic, M., Smith, N., et al. (2007). Restriction landmark genome scanning identifies culture-induced DNA methylation instability in the human embryonic stem cell epigenome. Hum. Mol. Genet. *16*, 1253–1268.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., et al. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947–956.

Brimble, S.N., Sherrer, E.S., Uhl, E.W., Wang, E., Kelly, S., Merrill, A.H., Jr., Robins, A.J., and Schulz, T.C. (2007). The cell surface glycosphingolipids SSEA-3 and SSEA-4 are not essential for human ESC pluripotency. Stem Cells *25*, 54–62.

Buecker, C., Chen, H.-H., Polo, J.M., Daheron, L., Bu, L., Barakat, T.S., Okwieka, P., Porter, A., Gribnau, J., Hochedlinger, K., et al. (2010). A murine ESC-like state facilitates transgenesis and homologous recombination in human pluripotent stem cells. Cell Stem Cell *6*, 535–546.

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. Nature *450*, 1230–1234.

Chen, G., Gulbranson, D.R., Hou, Z., Bolin, J.M., Ruotti, V., Probasco, M.D., Smuga-Otto, K., Howden, S.E., Diol, N.R., Propson, N.E., et al. (2011). Chemically defined conditions for human iPSC derivation and culture. Nat. Methods *8*, 424–429.

Chew, J.-L., Loh, Y.-H., Zhang, W., Chen, X., Tam, W.-L., Yeap, L.-S., Li, P., Ang, Y.-S., Lim, B., Robson, P., et al. (2005). Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells. Mol. Cell. Biol. *25*, 6031–6046.

Collier, A.J., Panula, S.P., Schell, J.P., Chovanec, P., Plaza Reyes, A., Petropoulos, S., Corcoran, A.E., Walker, R., Douagi, I., Lanner, F., et al. (2017). Comprehensive cell surface protein profiling identifies specific markers of human naive and primed pluripotent states. Cell Stem Cell *20*, 874–890.e7.

Crombie, D.E., Daniszewski, M., Liang, H.H., Kulkarni, T., Li, F., Lidgerwood, G.E., Conquest, A., Hernández, D., Hung, S.S., Gill, K.P., et al. (2017). Development of a modular automated system for maintenance and differentiation of adherent human pluripotent stem cells. SLAS Discov. *22*, 1016–1025.

Daniszewski, M., Crombie, D.E., Henderson, R., Liang, H.H., Wong, R.C.B., Hewitt, A.W., and Pébay, A. (2017). Automated cell culture systems and their applications to human pluripotent stem cell studies. SLAS Technol. *23*, 315–325.

Furue, M.K., Na, J., Jackson, J.P., Okamoto, T., Jones, M., Baker, D., Hata, R.-I., Moore, H.D., Sato, J.D., and Andrews, P.W. (2008). Heparin promotes the growth of human embryonic stem cells in a defined serum-free medium. Proc. Natl. Acad. Sci. USA *105*, 13409–13414.

Guo, G., and Smith, A. (2010). A genome-wide screen in EpiSCs identifies Nr5a nuclear receptors as potent inducers of ground state pluripotency. Development *137*, 3185–3192.

Hanna, J., Markoulaki, S., Mitalipova, M., Cheng, A.W., Cassady, J.P., Staerk, J., Carey, B.W., Lengner, C.J., Foreman, R., Love, J., et al. (2009). Metastable pluripotent states in NOD-mouse-derived ESCs. Cell Stem Cell *4*, 513–524.

Hough, S.R., Thornton, M., Mason, E., Mar, J.C., Wells, C.A., and Pera, M.F. (2014). Single-cell gene expression profiles define self-renewing, pluripotent, and lineage primed states of human pluripotent stem cells. Stem Cell Rep. *2*, 881–895.

Hussein, S.M., Batada, N.N., Vuoristo, S., Ching, R.W., Autio, R., Närvä, E., Ng, S., Sourour, M., Hämäläinen, R., Olsson, C., et al. (2011). Copy number variation and selection during reprogramming to pluripotency. Nature *471*, 58–62.

International Stem Cell Initiative Consortium, Akopian, V., Andrews, P.W., Beil, S., Benvenisty, N., Brehm, J., Christie, M., Ford, A., Fox, V., Gokhale, P.J., et al. (2010). Comparison of defined culture systems for feeder cell free propagation of human embryonic stem cells. In Vitro Cell. Dev. Biol. Anim. *46*, 247–258.

Kilpinen, H., Helena, K., Angela, G., Andreas, L., Vackar, A., Sofie, A., Sendu, B., Dalila, B., Casale, F.P., Oliver, C., et al. (2016). Common genetic variation drives molecular heterogeneity in human iPSCs. Nature *546*, 370–375.

Kim, M.-H., and Kino-oka, M. (2018). Bioprocessing strategies for pluripotent stem cells based on Waddington's epigenetic landscape. Trends Biotechnol. *36*, 89–104.

Klimanskaya, I., Chung, Y., Meisner, L., Johnson, J., West, M.D., and Lanza, R. (2005). Human embryonic stem cells derived without feeder cells. Lancet *365*, 1636–1641.

Konagaya, S., Ando, T., Yamauchi, T., Suemori, H., and Iwata, H. (2015). Long-term maintenance of human induced pluripotent stem cells by automated cell culture system. Sci. Rep. *5*, 16647.

Lin, P., Troup, M., and Ho, J.W.K. (2017). CIDR: ultrafast and accurate clustering through imputation for single-cell RNA-seq data. Genome Biol. *18*, 59.

Loh, Y.-H., Wu, Q., Chew, J.-L., Vega, V.B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., et al. (2006). The Oct4 and nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat. Genet. *38*, 431–440.

Ludwig, T.E., Levenstein, M.E., Jones, J.M., Berggren, W.T., Mitchen, E.R., Frane, J.L., Crandall, L.J., Daigh, C.A., Conard, K.R., Piekarczyk, M.S., et al. (2006). Derivation of human embryonic stem cells in defined conditions. Nat. Biotechnol. *24*, 185–187.

Najm, F.J., Chenoweth, J.G., Anderson, P.D., Nadeau, J.H., Redline, R.W., McKay, R.D.G., and Tesar, P.J. (2011). Isolation of epiblast stem cells from preimplantation mouse embryos. Cell Stem Cell *8*, 318–325.

Ng, H.-H., and Surani, M.A. (2011). The transcriptional and signalling networks of pluripotency. Nat. Cell Biol. *13*, 490–496.

Nguyen, Q., Lukowski, S., Chiu, H., Friedman, C., Senabouth, A., Crowhurst, L., Bruxmer, T., Christ, A., Palpant, N., and Powell, J. (2017). Determining cell fate specification and genetic contribution to cardiac disease risk in hiPSC-derived cardiomyocytes at single cell resolution. bioRxiv. https://doi.org/10.1101/2293.

Nguyen, Q.H., Lukowski, S.W., Chiu, H.S., Senabouth, A., Bruxner, T.J.C., Christ, A.N., Palpant, N.J., and Powell, J.E. (2018). Single-cell RNA-seq of human induced pluripotent stem cells reveals cellular heterogeneity and cell state transitions between subpopulations. Genome Res. *28*, 1053–1066.

O'Brien, C.M., Chy, H.S., Zhou, Q., Blumenfeld, S., Lambshead, J.W., Liu, X., Kie, J., Capaldo, B.D., Chung, T.-L., Adams, T.E., et al. (2017). New monoclonal antibodies to defined cell surface proteins on human pluripotent stem cells. Stem Cells *35*, 626–640.

Park, I.-H., Zhao, R., West, J.A., Yabuuchi, A., Huo, H., Ince, T.A., Lerou, P.H., William Lensch, M., and Daley, G.Q. (2007). Reprogramming of human somatic cells to pluripotency with defined factors. Nature *451*, 141–146.

Paull, D., Sevilla, A., Zhou, H., Hahn, A.K., Kim, H., Napolitano, C., Tsankov, A., Shang, L., Krumholz, K., Jagadeesan, P., et al. (2015). Automated, high-throughput derivation, characterization and differentiation of induced pluripotent stem cells. Nat. Methods *12*, 885–892.

Pébay, A., Wong, R.C.B., Pitson, S.M., Wolvetang, E.J., Peh, G.S.-L., Filipczyk, A., Koh, K.L.L., Tellis, I., Nguyen, L.T.V., and Pera, M.F. (2005). Essential roles of sphingosine-1-phosphate and platelet-derived growth factor in the maintenance of

human embryonic stem cells. Stem Cells *23*, 1541–1548.

Radzisheuskaya, A., Chia, G.L.B., dos Santos, R.L., Theunissen, T.W., Castro, L.F.C., Nichols, J., and Silva, J.C.R. (2013). A defined Oct4 level governs cell state transitions of pluripotency entry and differentiation into all embryonic lineages. Nat. Cell Biol. *15*, 579–590.

Reubinoff, B.E., Pera, M.F., Fong, C.Y., Trounson, A., and Bongso, A. (2000). Embryonic stem cell lines from human blastocysts: somatic differentiation in vitro. Nat. Biotechnol. *18*, 399–404.

Rizzino, A. (2009). Sox2 and Oct-3/4: a versatile pair of master regulators that orchestrate the self-renewal and pluripotency of embryonic stem cells. Wiley Interdiscip. Rev. Syst. Biol. Med. *1*, 228–236.

Sun, N., Panetta, N.J., Gupta, D.M., Wilson, K.D., Lee, A., Jia, F., Hu, S., Cherry, A.M., Robbins, R.C., Longaker, M.T., et al. (2009). Feeder-free derivation of induced pluripotent stem cells from adult human adipose stem cells. Proc. Natl. Acad. Sci. USA *106*, 15720–15725.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663–676.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell *131*, 861–872.

Thomson, J.A. (1998). Embryonic stem cell lines derived from human blastocysts. Science *282*, 1145–1147.

Vallier, L., Alexander, M., and Pedersen, R.A. (2005). Activin/nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. J. Cell Sci. *118*, 4495–4509.

Warrier, S., Van der Jeught, M., Duggal, G., Tilleman, L., Sutherland, E., Taelman, J., Popovic, M., Lierman, S., Chuva De Sousa Lopes, S., Van Soom, A., et al. (2017). Direct comparison of distinct naive pluripotent states in human embryonic stem cells. Nat. Commun. *8*, 15055.

Weinberger, L., Ayyash, M., Novershtern, N., and Hanna, J.H. (2016). Dynamic stem cell states: naive to primed pluripotency in rodents and humans. Nat. Rev. Mol. Cell Biol. *17*, 155–169.

Xu, R.-H., Peck, R.M., Li, D.S., Feng, X., Ludwig, T., and Thomson, J.A. (2005). Basic FGF and suppression of BMP signaling sustain undifferentiated proliferation of human ES cells. Nat. Methods *2*, 185–190.

Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., et al. (2007). Induced pluripotent stem cell lines derived from human somatic cells. Science *318*, 1917–1920.

Yu, P., Nie, Q., Tang, C., and Zhang, L. (2018). Nanog induced intermediate state in regulating stem cell differentiation and reprogramming. BMC Syst. Biol. *12*, 22.

ISCI, Volume *7*

Supplemental Information

Single-Cell Profiling Identifies Key Pathways Expressed by iPSCs Cultured in Different Commercial Media

Maciej Daniszewski, Quan Nguyen, Hun S. Chy, Vikrant Singh, Duncan E. Crombie, Tejal Kulkarni, Helena H. Liang, Priyadharshini Sivakumaran, Grace E. Lidgerwood, Damián Hernández, Alison Conquest, Louise A. Rooney, Sophie Chevalier, Stacey B. Andersen, Anne Senabouth, James C. Vickers, David A. Mackey, Jamie E. Craig, Andrew L. Laslett, Alex W. Hewitt, Joseph E. Powell, and Alice Pébay

Supplementary Figures and Legends

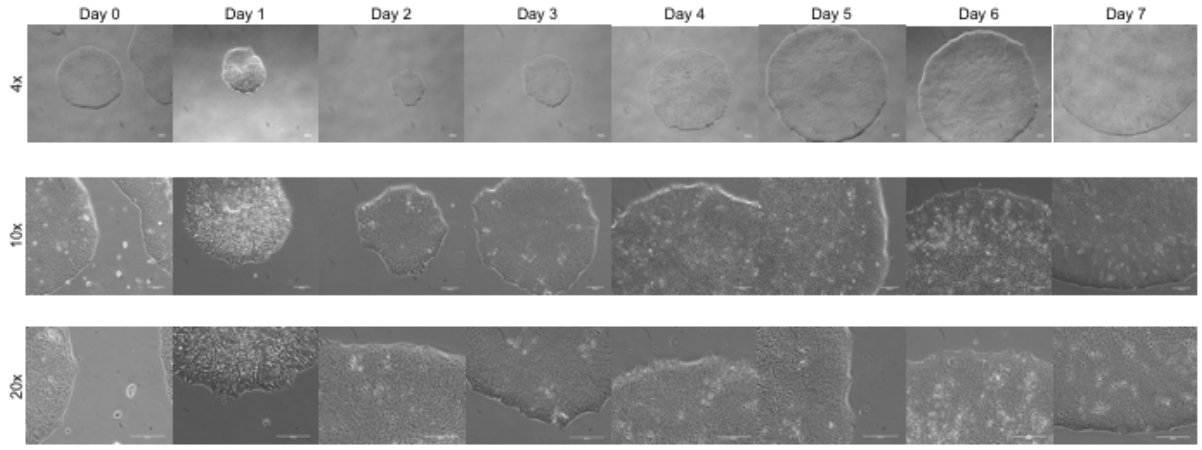


Figure S1: Representative morphology of iPSCs maintained in StemFlex™, Related to Figure 1. Images of a human iPSC line (MBE2906), from day 0 to day 7, at different magnifications (x4, 10 and 20). Cells were cultivated using an automated platform, on Vitronectin and in StemFlex™; medium was changed every 2-3 days. Images are representative of all cell lines. Scale bars: 100 µM.

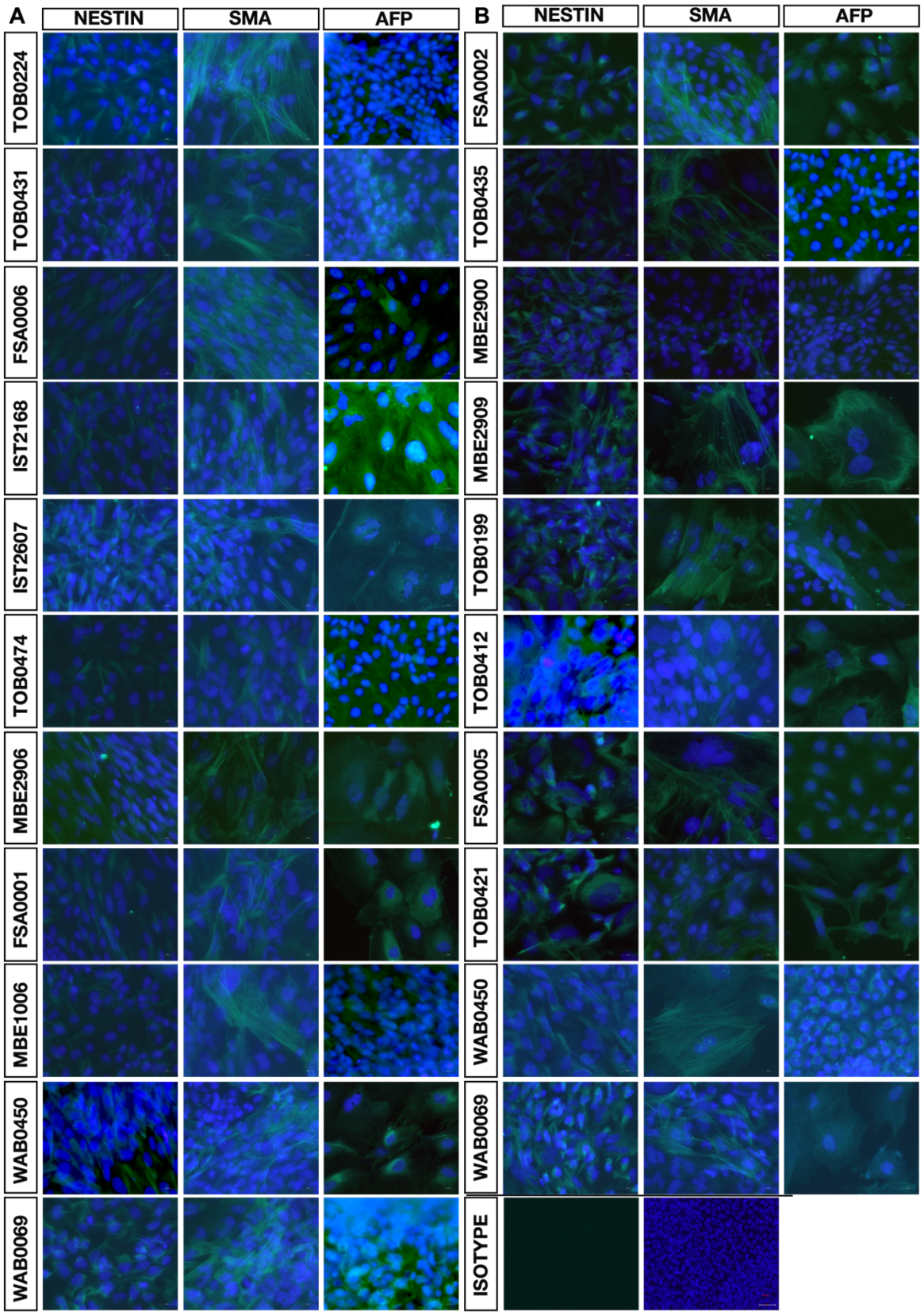


Figure S2: Differentiation of iPSCs into the three germ layers, Related to Figures 1, 2 and 3. Embryoid bodies of iPSCs maintained in (**A**) StemFlex™ and (**B**) TeSR™-E8™ showing differentiation into the three germ layers by immunostaining with NESTIN, SMA, AFP and with DAPI nucleic acid counterstain. Bottom right: representative negative isotype control with its corresponding DAPI. Scale bars: 20 μM.

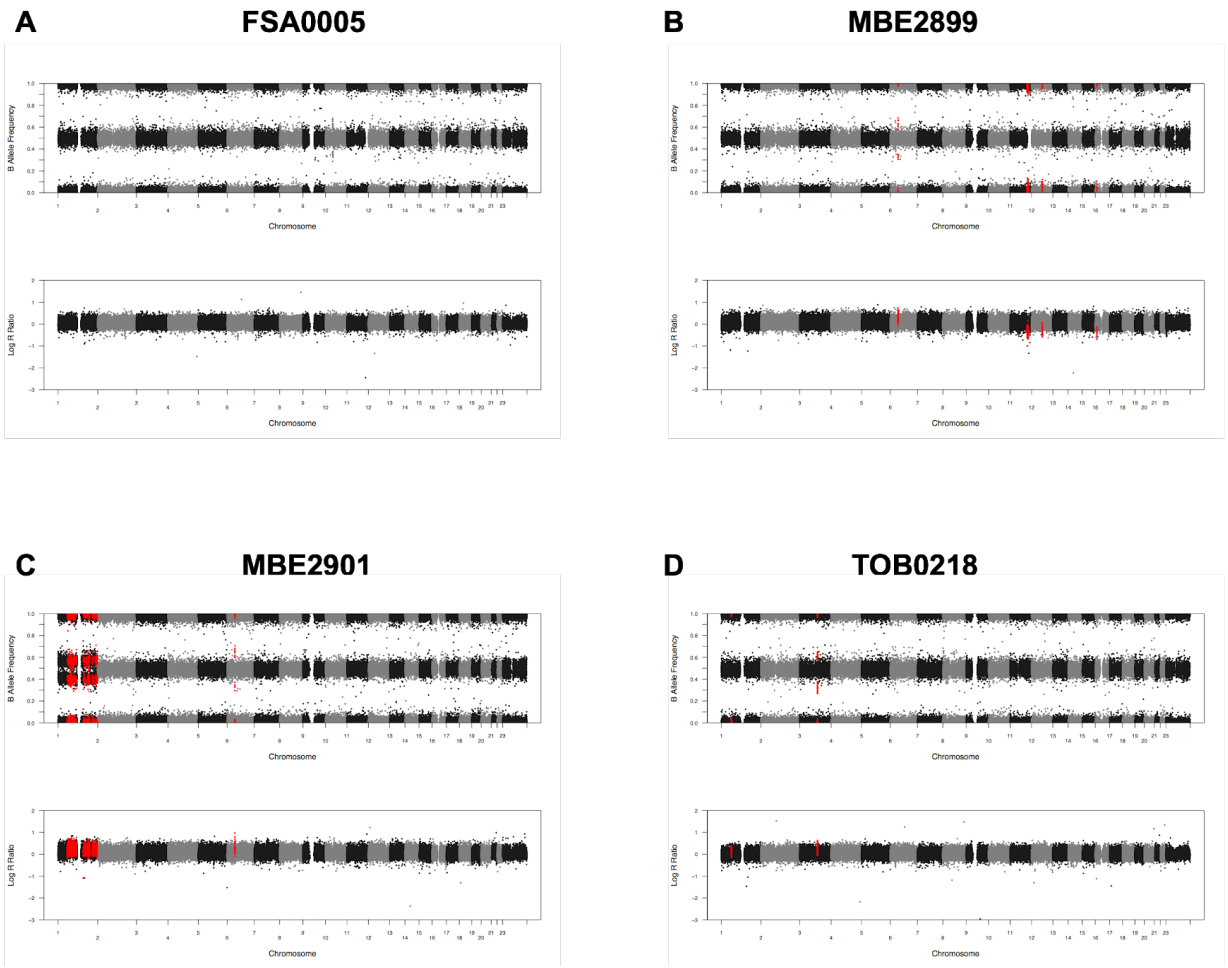


Figure S3: Copy Number Variation Analysis, Related to Figures 1 and 2. (**A**) Representative analysis of iPSCs with a normal virtual karyotyping (FSA0005). Anomalies were revealed in the iPSC lines MBE2899 (**B**), MBE2901 (**C**) and TOB0218 (**D**). Each panel shows the B allele frequency (BAF) and the log R ratio (LRR). BAF at values others then 0, 0.5 or 1 indicate an abnormal copy number. Similarly, the LRR represents a logged ratio of "observed probe intensity to expected intensity". A deviation from zero corresponds to a change in copy number.

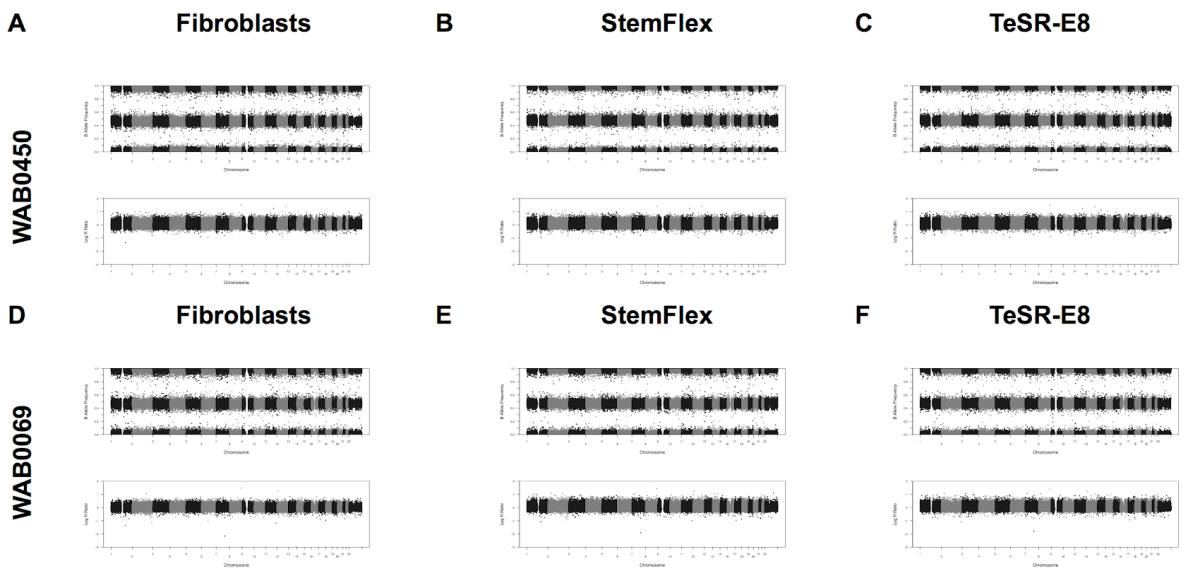


Figure S4: Characterisation of WAB0450 and WAB0069, Related to Figure 3. (**A-F**) Copy Number Variation Analysis of WAB0450 (**A-C**) and WAB0069 (**D-F**) in original fibroblasts (**A, D**), iPSCs at p8 in StemFlex™ (**B, E**) and TeSR™-E8™ (**C, F**). Each panel shows the B allele frequency (BAF) and the log R ratio (LRR).

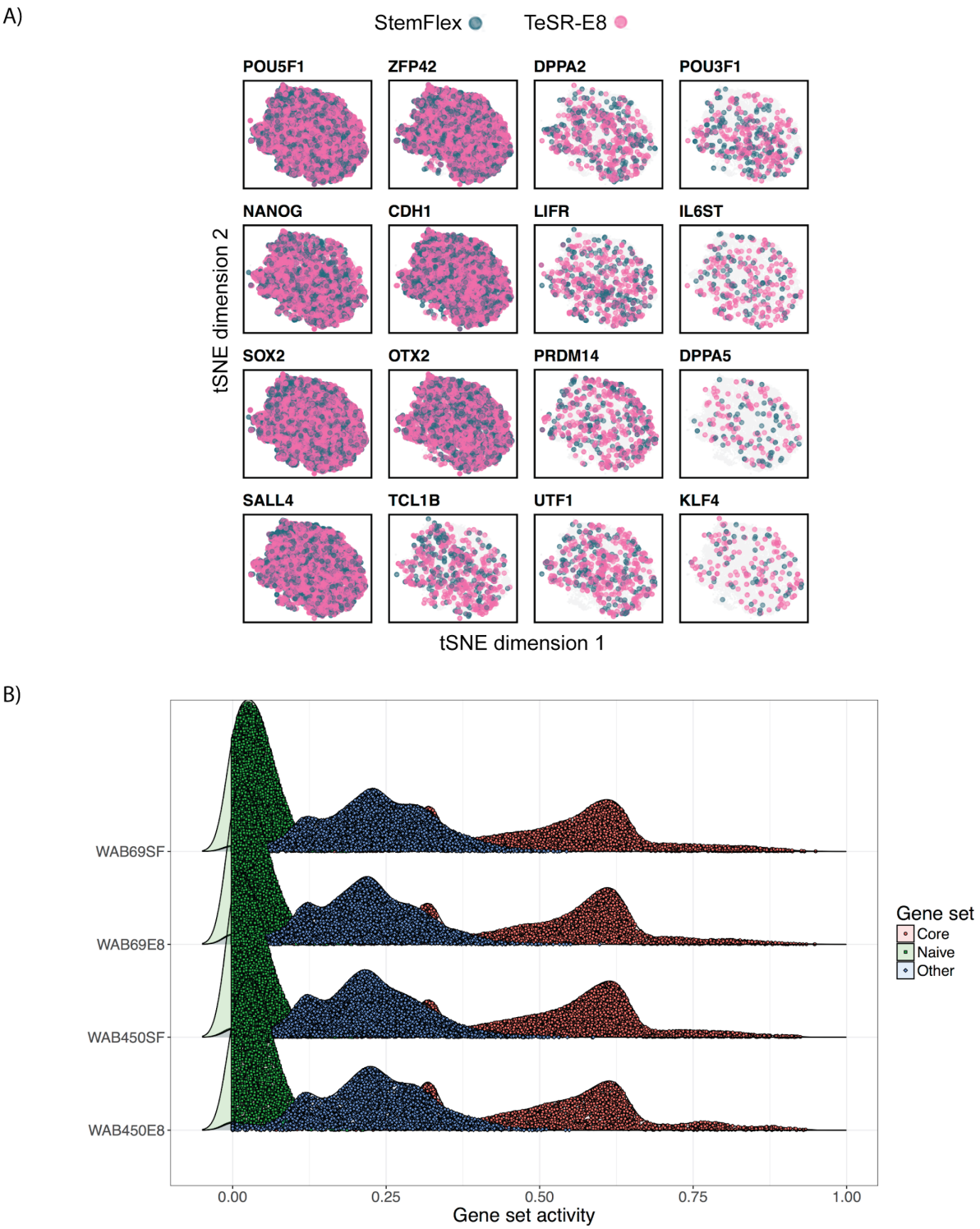


Figure S5. Expression of key pluripotency gene markers in the two media, Related to Figure 3 and Tables S1-2. (A) Cells in StemFlex™ (the pool of WAB0450SF and WAB0069SF cells) and TeSR™-E8™ (the pool of WAB0450E8 and WAB0069E8) expressing pluripotency markers are displayed in two dimensional tSNE plots. The majority of cells express common key pluripotency markers, such as *POU5F1* (*OCT4*), NANOG, *SOX2*, and *OTX2*. For all 16 genes shown in the figure, the number of expressing cells are even between the two media. (**B)** Gene set activities calculated by applying AUCell approach for three sets of known markers shown in Table S2. The regulation activity, ranging from 0 to 1, is the enrichment of the genes in the input gene set across the expression ranking of all genes in a cell. The three sets include "Core pluripotency markers" (Core), "Naïve pluripotency markers" (Naïve), and "Other pluripotency markers" (Other).

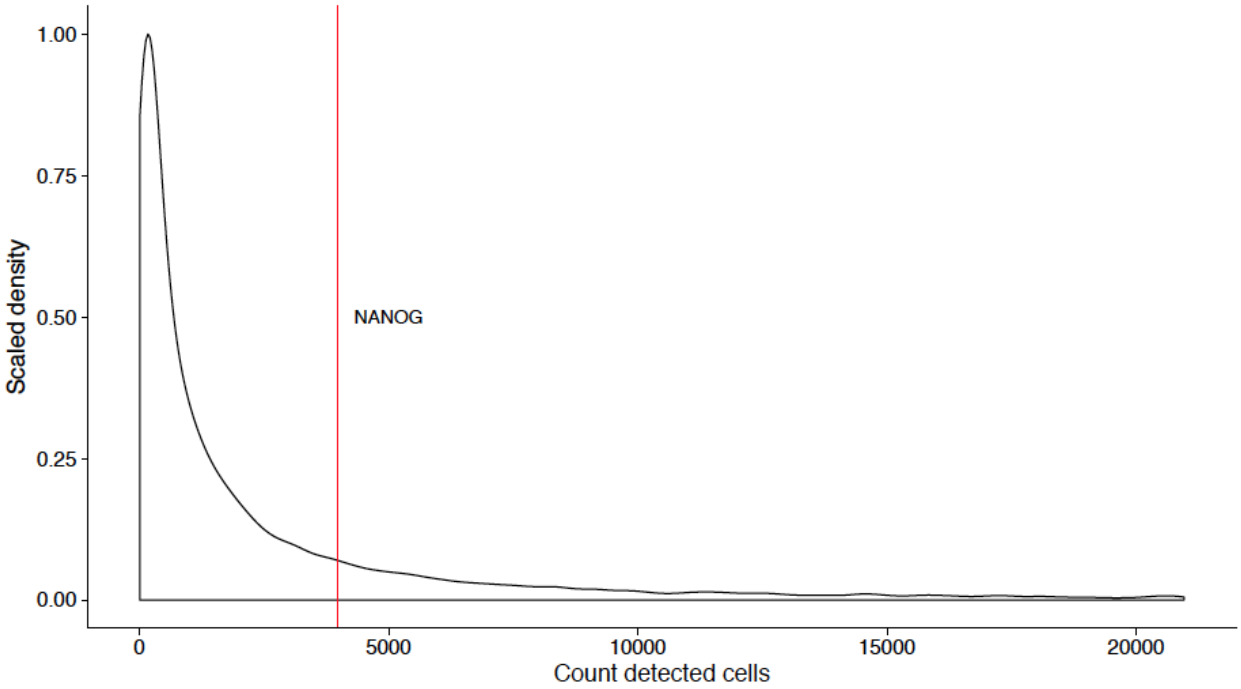


Figure S6. Distribution of *NANOG* in all cells, Related to Figure 3 and Tables S1-2. The percentage of cells expressing *NANOG* is at 75.5th percentile of all 16,270 reliably detected genes, higher than 13,349 genes.

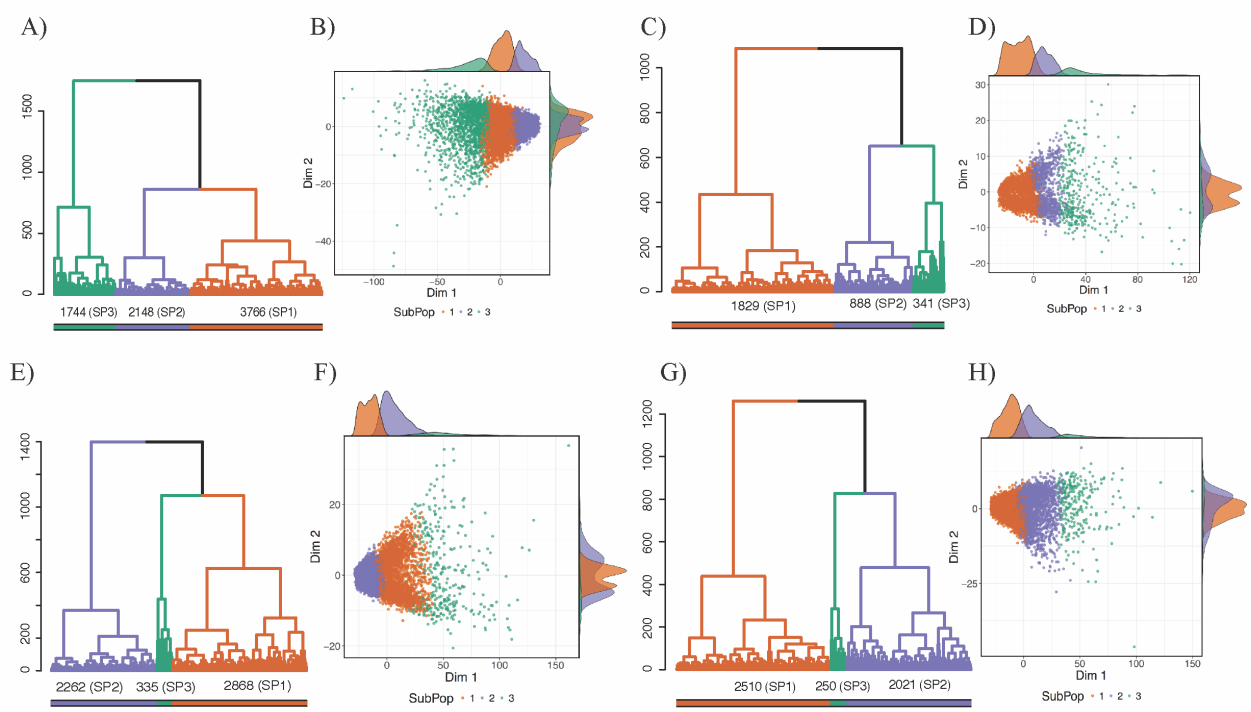


Figure S7: Single-cell subpopulation analysis of WAB0450 and WAB0069 in StemFlex™ and TeSR™-E8™ media, Related to Figure 3 and Tables S3-6. Subpopulations were identified by our unsupervised clustering algorithm so that a subpopulation consist of cells that are more similar to cells within the subpopulation compared to cells in other subpopulations. For each sample, a dendrogram tree with branches colored by subpopulations and the numbers of cells in each subpopulation are shown (panels **A**, **C**, **E**, and **G** for WAB0450SF, WAB0450E8, WAB0069SF and WAB0069E8, respectively). The distribution of cells in two first PCA principal components are shown as scatter plots accompanied by density plots on the right and top axes (panels **B**, **D**, **F**, and **H** for WAB0450SF, WAB0450E8, WAB0069SF and WAB0069E8 respectively).

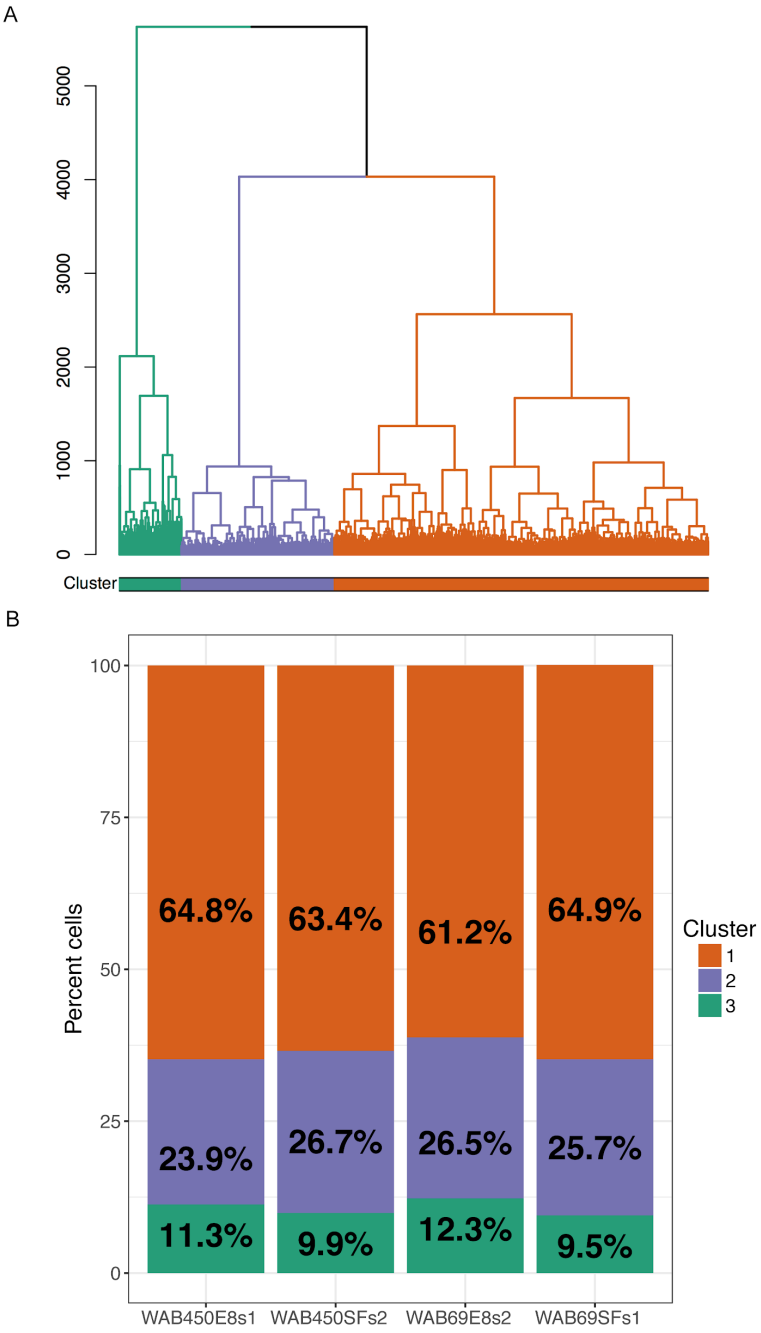


Figure S8. Subpopulation analysis of all cells in all four samples, Related to Figure 3 and Tables S3-6. (**A**) Dendrogram showing clustering results for more than 20,000 cells from the merged expression data of all four samples. The three clusters are shown by different colored branches. (**B**) The percent of cells distributed to each of the three clusters for each sample.

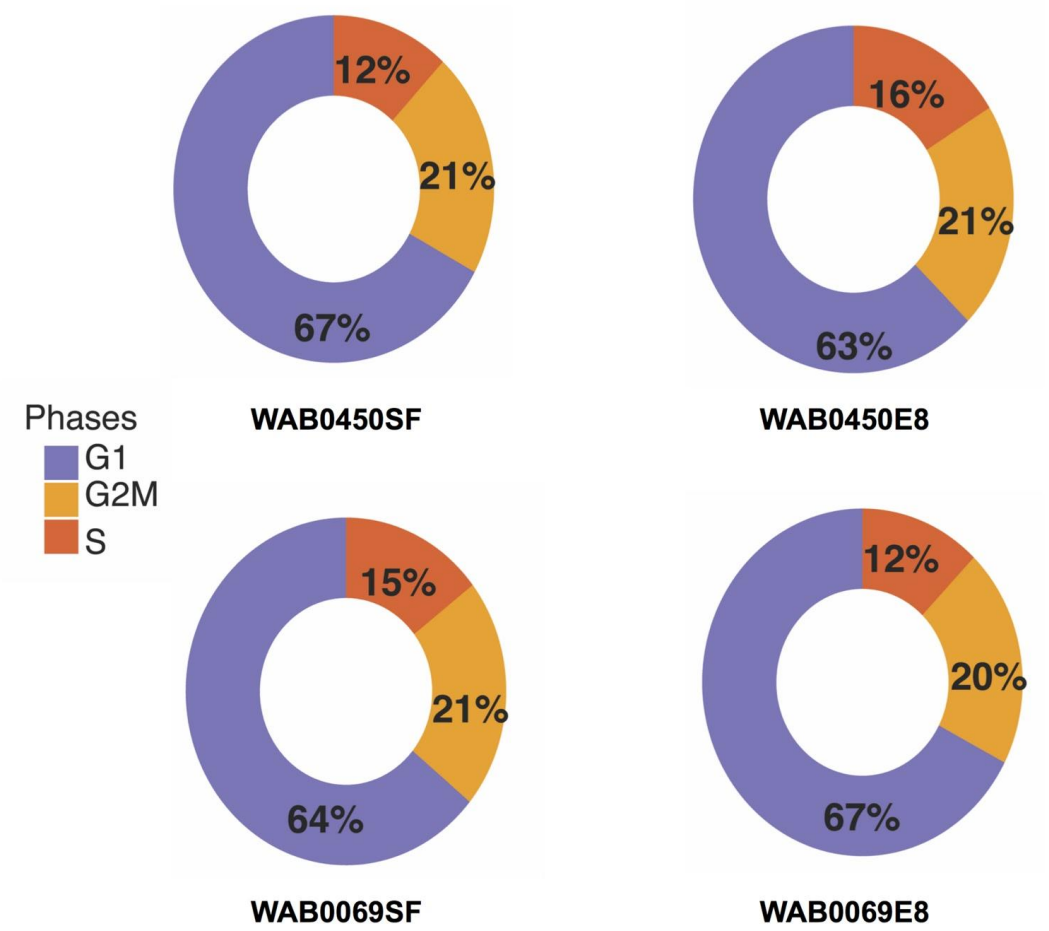


Figure S9. Predicted cell cycle stage for every cell in all combinations, Related to Figure 3 and Tables S3-6. WAB0450 and WAB0069 cells in StemFlex (SF) and TeSR-E8 (E8) showed no significant difference in proportions in the different cell cycle phases (G1, G2M, S).

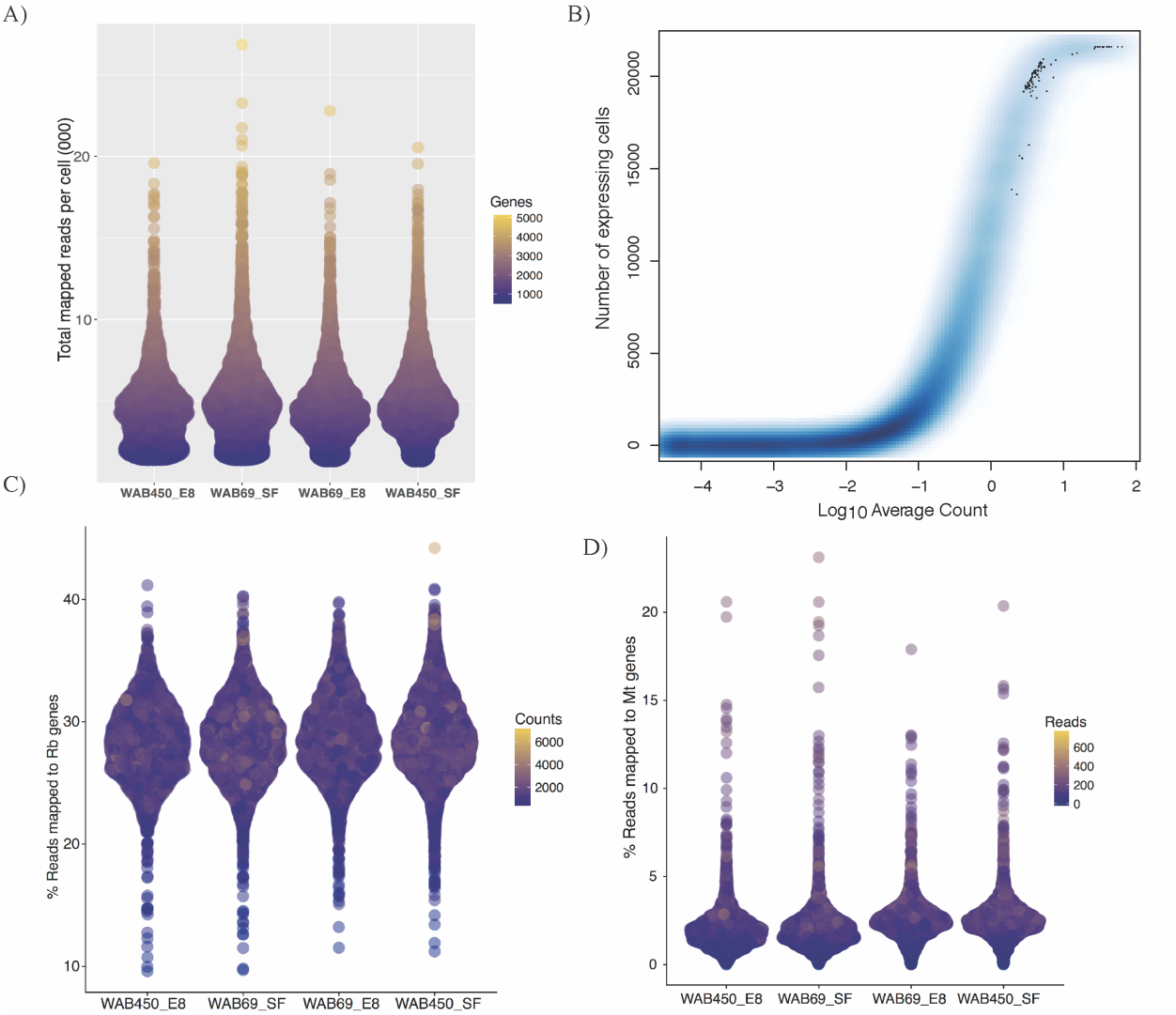


Figure S10: Quality control statistics for the single cell RNA sequencing of 21,597 human iPSCs, Related to Figure 3. (**A**)The library size (total mapped reads) range for each cell in each of the four samples. (**B**) Expression range of all genes detected. The range is greater than 6 orders of magnitudes. Genes detected in fewer than 0.1% of all cells were removed. (**C**) Percent reads mapped to mitochondrial genes as an indicator of cell conditions. The *x*-axis shows four samples, and the y-axis shows reads mapped. Cells with mitochondrial reads exceeding 3 x median absolute deviation (MAD) or mitochondrial reads above 20% were removed. (**D**) Percent reads mapped to ribosomal genes. We removed cells with over 3 x MAD or over 50% reads mapped to ribosomal genes.

Supplemental Tables added as Supplemental Data

Table S1. Reactome and DE genes. Differential expression and functional enrichment analysis (Reactome differences in SF and E8; WAB0069 reactome differences; WAB0450 reactome differences; DE genes in SF and E8_2lines), Related to Figures 3, S5 and S6.

Table S2. Comparative summary of levels of expression of key genes involved in pluripotency and metabolism in StemFlex™ and TeSR™-E8™ media (pluripotency; proliferation and survival; metabolism), Related to Figures 3, S5 and S6.

Table S3. Pathway analysis subpopulations, Related to Figures 3, S7-S9.

Table S4. Comparing subpopulations between four samples (same cell line or different cell lines) by a machine-learning classification approach, Related to Figures 3, S7-S9.

Table S5. Percent positive cells by subpopulations, Related to Figures 3, S7-S9.

Table S6. Pathway Enrichment & Gene Markers, Related to Figures 3, S7-S9.

Transparent Methods

Ethics. All experimental work performed in this study was approved by the Human Research Ethics committees of the Royal Victorian Eye and Ear Hospital (11/1031H) and University of Tasmania (H0014124) with the requirements of the National Health & Medical Research Council of Australia (NHMRC) and conformed with the Declarations of Helsinki (McCaughey et al., 2016).

Fibroblast Culture. Human skin biopsies were obtained from subjects over the age of 18 years. Fibroblasts were cultured in DMEM with high glucose, 10% fetal bovine serum (FBS), L-glutamine, 100 U/mL penicillin and 100 µg/mL streptomycin (all from Thermo Fisher Scientific, USA). All cell lines were mycoplasma-free (MycoAlert mycoplasma detection kit, Lonza, Switzerland). Fibroblasts at passage (p) 8-10 were used for reprogramming.

Generation, selection and maintenance of iPSCs. A TECAN liquid handling platform was used to cultivate cells, as described in (Crombie et al., 2017). iPSCs were generated by nucleofection (Amaxa™ Nucleofector™) with episomal vectors expressing *OCT-4*, *SOX2*, *KLF4*, *L-MYC*, *LIN28* and shRNA against *p53* (Okita et al., 2011) in feeder- and serum-free conditions using TeSR™-E7™ medium (Stem Cell Technologies, Canada) as described previously (Crombie et al., 2017). The reprogrammed cells were maintained on the automated platform using TeSR™-E7™ medium, with daily medium change. Pluripotent cells were selected by sorting anti-human TRA-1-60 Microbeads using a MultiMACS (Miltenyi, Germany) as described in (Crombie et al., 2017). Cell number was determined and cell were subsequently plated onto vitronectin XF™ (Stem Cell Technologies) in TeSR™-E8™ medium (Stem Cell Technologies) or StemFlex™ medium (Thermo Fisher Scientific). Subsequent culturing was performed on the automated platform using TeSR™-E8™ or StemFlex™ medium, and medium was changed every two days. Passaging was performed weekly on the automated platform using ReLeSR™ (Stem Cell Technologies) onto vitronectin XF™ coated plates as described in (Crombie et al., 2017)**.**

iPSC quality control. TRA-1-60 quantifications were performed on a MACSQUANT immediately prior to passaging to fresh plates as described (Crombie et al., 2017). Pluripotency was assessed by expression of the markers OCT3/4 (sc-5279, Santa Cruz Biotechnology, USA), TRA-1-60 (MA1-023-PE, Thermo Fisher Scientific, USA). Copy number variation (CNV) analysis of original fibroblasts and iPSCs was performed using Illumina HumanCore Beadchip arrays. CNV analyses were performed using PennCNV (Colella et al., 2007) and QuantiSNP (Wang et al., 2007) with default parameter settings. Chromosomal aberrations were deemed to involve at least 20 contiguous single nucleotide polymorphisms (SNPs) or a genomic region spanning at least 1MB (Colella et al., 2007)(Wang et al., 2007). The B allele frequency (BAF) and the log R ratio (LRR) were extracted from GenomeStudio (Illumina, USA) for representation. Day 14 embryoid bodies (EBs) were plated on 24-well plates coated with 0.1% gelatin to differentiate for 7 days cultured in differentiation medium: DMEM, 20% FBS, 1% Non-Essential Amino Acid solution, 0.1mM β-mercaptoethanol and 1% Pen-strep. Germ layer differentiation was assessed by immunochemistry. For immunochemistry, cells were fixed with 4% paraformaldehyde, permeabilized with 0.1% Triton X-100 (Sigma-Aldrich) and blocked with 2% normal goat serum (Life Technologies). Standard immunostaining procedures were performed using the following primary antibodies: mouse anti- Nestin (Abcam, AB22035) for ectoderm; mouse anti- Smooth Muscle Actin (SMA, clone 1A4, R&D Systems, MAB1420) for mesoderm; mouse anti- Alpha-fetoprotein (AFP, Merck Millipore, ST1673) for endoderm. Secondary antibody used was Alexa Fluor 488 goat anti-mouse IgG (ThermoFisher, A11029). Nuclei were counterstained with DAPI (Sigma Aldrich).

Flow cytometry. The intracellular and extracellular immunolabelling of iPSCs and FACS analyses were all similarly performed as previously described (O'Brien et al., 2017). Briefly, cells grown in culture flasks or dishes were washed twice with 5 mL PBS -/- (Gibco, Life Technologies) and harvested using 1-2ml TrypLE Express (Gibco, Life Technologies) incubated for 5 minutes at 37°C. The cells were washed twice with PBS supplemented with 10% v/v FBS, which neutralises the TrypLE™ Express and help with cell viability before staining. The cells were resuspended in 10% v/v PBS -/- and aliquoted into individual 5ml FACs tubes for single extracellular labelling with primary monoclonal antibodies Anti-GPR64 (CSIRO, CSTEM7), Anti-CDCP1 (CSIRO, CSTEM26), Anti-F11R (CSIRO, CSTEM27), Anti-DSG2 (CSIRO, CSTEM28), Anti-CDH3 (CSIRO, CSTEM29), Anti-NLGN4X (CSIRO, CSTEM30), and Anti-PCDH1 (CSIRO, CSTEM31) and the stem cell markers TRA-1-60 and SSEA3 (both Merck Millipore). Primary and secondary antibodies were incubated for 20 minutes on ice and washed twice with PBS. The fluorescent secondary antibody goat anti-mouse IgG Alexa Fluor™ 488 (AF488, Thermo Fisher Scientific) was used to counterstain the primary antibodies. Non-cross reacted secondary antibodies were washed away with PBS and cells were resuspended with PBS containing 0.1% v/v propidium iodide (Sigma-Aldrich, USA). The propidium iodide was used to exclude non-viable cells during FACS analysis. For intracellular staining with Anti-OCT3/4 (Merck Millipore), cells were fixed with 4% paraformaldehyde for 30 minutes at RT, permeabilised with 0.01% Triton-X100 (Sigma-Aldrich) for 5 minutes and blocked with 10% Goat Serum (Thermo Fisher Scientific) for 30 minutes at RT after two washes with PBS. The fixed cells were labelled with Anti-OCT3/4 for 20 minutes at RT and washed twice with PBS. The secondary conjugated goat anti-mouse IgG1 AF488 fluorophore (Thermo Fisher Scientific) was used for the cross-reaction with Anti-OCT3/4 antibodies. The washed cells were finally resuspended in PBS for FACS analysis using BD™ Biosciences LSR II cell analyzer.

Single cell RNA sequencing. Cells were harvested using ReleSR™, colonies were dissociated into single cell suspension. Cells were counted and assessed for viability with Trypan Blue using a Countess II automated counter (Thermo Fisher Scientific), then pooled at a concentration of 391-663 cells/µL (3.91x10^5-6.63x10^5 cells/mL). Final cell viability estimates ranged between 85-97%. Cells were partitioned and barcoded using high-throughput droplet 10X Genomics Chromium™ Controller (10X Genomics, USA) and the Single Cell 3' Library and Gel Bead Kit (V2; 10X Genomics; PN-120237). The estimated number of cells in each well in the Chromium™ chip was optimized to capture approximately 8000-10,000 cells per cell pool. GEM generation and barcoding, cDNA amplification, and library construction were performed according to standard protocol. The resulting single cell transcriptome libraries were pooled and sequenced on an Illumina NextSeq™500, using a 150 cycle High Output reagent kit (NextSeq™500/550 v2; Illumina) in standalone mode as follows: 26bp (Read 1), 8bp (Index), and 98 bp (Read 2).

Bioinformatics mapping of reads to original genes and cells. Raw sequencing data (Illumina BCL files) were processed directly with the *cellranger* pipeline v2.0.0 (*mkfastq*, *count*, *aggr*) using the default parameters, except for the estimated cell number set at 5,000 cells. The reads were aligned to the *homo sapiens* GRCh38p10 Cell Ranger 2.0 reference genome using the STAR software (Dobin et al., 2013). Cell barcodes and unique molecular identifiers (UMI) were filtered using default parameters in the *cellranger count* processing. Reads with > 90% base calling accuracy were kept and UMI sequences were corrected based on high-quality sequenced UMIs with Hamming distance 1. A gene count expression matrix for four samples was generated using the *cellranger aggr* function.We have made available both the raw and normalised data on ArrayExpress under accession E-MTAB-6524.

Combining individual and cell SNP genotypes to assign cells to samples. We developed a new computational pipeline to combine SNP-chip data with scRNA SNP calling data in

individual cell to de-multiplex mixture of cells from different samples pooled into one sequencing reaction. Two cell lines, including WAB0450 and WAB0069 in TeSR™-E8™ and StemFlex™ media, were genotyped separately by Infinium HumanCore-24 v1.1 BeadChip assay (Illumina). GenomeStudio™ V2.0 (Illumina) was used for SNP genotype calling of the BeadChip data. The full genotype report files were reformatted into Plink map, fam, and lgen files using custom Shell script and were then converted into variant calling format (vcf) using Plink 2 (Chang et al., 2015). For each sorted, indexed vcf file (separated by chromosomes), a strand fixing step was performed using bcf fixref function (Li, 2011). To increase the genome coverage of the BeadChip genotype data (total 306,670 SNPs), we performed imputation to the whole genome. Prior to imputation, Eagle V.2.3.5 was used for haplotype phasing the strand-fixed genotype vcf files (Loh et al., 2016). The phased data were imputed based on the 1000 genome phase 3 reference panel (2,535 samples) using the *minimac3* program (Fuchsberger, C Abecasis, GR Hinds, DA, n.d.). Single-cell SNPs were called from the mapped RNA BAM files using Freebayes V1.0.2 (Garrison and Marth, 2012). The likelihood that a cell originated from a sample is the cumulative likelihoods of variants identified in each cell. We applied Demuxlet software (Kang et al., 2017) to calculate posterior probability of a genotype *g* identified for a cell based on scRNA data given the DNA data from the imputed BeadChip genotypes. Each cell was assigned to the sample with the highest likelihood.

scRNA data filtering and normalisation. Sequencing reads from four samples were pooled and normalised at two levels - between samples and between cells. Briefly, to account for potential systematic measurement variation in total read depth per sample, we utilised unique information in read unique molecular identifiers (UMIs), sample indexes, and cell barcodes to perform random binomial sampling of reads. From all reads mapped to each gene in each cell, a subset of reads was randomly drawn from the binomial distribution, at the rate $Rate_i$, ($Reads_g$, $Rate_i$). $Rate_i$ was determined based on: total reads per sample, number of cells per sample, and fraction of mapped reads in a sample. This method has been shown to be less biased than other global-scaling methods and can equalise between sample sequencing depth while maintaining original read distribution (Zheng et al., 2017). The total mapped reads per cell may be affected by sequencing artefacts such as dropouts (RNA molecules not amplified at the initial PCR amplification round), PCR duplicates, or high amount of ribosomal RNA. Cells with high proportion of reads mapped to mitochondrial are likely cells suffering from more stress due to technical cell handling during library preparation and cell sorting. Transcriptional changes in those cells, therefore, do not reflect the biological changes of interest. After sample aggregation and between sample normalisation, cells and genes were filtered to remove potential technical noise in the data, ensuring subsequent analysis was not affected by cells that were inconsistently outside the threshold of three times the Median Absolute Deviation (MAD), which is a more robust method to detect outlier than using standard deviation around the mean (Leys et al., 2013). We calculated MAD for total mapped reads per cell, total genes detected in a cell, expression of mitochondrial, and expression of ribosomal genes for every cell in the data set (Figure S10). Using the outside 3 x MAD range, we found 163 cells with low or high total mapped reads, and 456 cells by high mitochondrial gene expression and 57 cells high ribosomal genes. After removing cells, we subsequently removed 16,459 genes that were detected in less than 0.1% of all remaining cells (Figure S9B). Following sample normalisation and cell-gene filtering, a pooling and deconvolution method was implemented to normalise variation in read depth between cells (Lun et al., 2016). The method sequentially and randomly pools 40, 60, 80 and 100 cells to form pseudo cells, in which the expression values of a gene across cells are summed. A size factor was calculated for each pool based on the summed expression values. The pool size factors then deconstructed into the size factors of individual cells by QR decomposition. This approach alleviates the statistical challenge in a dataset with an inflation of zero count measures,

including stochastic zero expression of genes that are lowly expressed (higher dropout rates), or genes that are turned on/off in different subpopulations of cells.

Differential expression and functional enrichment analysis. We compared gene expression for cells in TeSR™-E8™ and StemFlex™ media in three experimental designs: the pool of two cell lines, WAB0069 alone, and WAB0450 alone. General linear model and negative binomial test were implemented to identify differentially expressed (DE) genes as described in the DESeq package (Anders and Huber, 2010). In each condition, each cell was considered as one biological replicate. Prior to estimating dispersion and normalization factors, a pseudo count of one was added to the UMI count expression matrix, which was subtracted post DE test to adjust for estimated fold-change. Significantly differentiated genes were those with *p*-adjusted values less than 5% (Benjamini-Hochberg adjustment for false discovery rate). To identify functional pathways different between cell lines and media, we applied Reactome functional network analysis, a reliably curated functional network (Wu et al., 2014), to find enriched pathways in DE gene lists. We predicted cell cycle stage for every cell, and found no differences in cell cycle effects in all conditions (Figure S9).Therefore, we performed normalisation between cells and opted not to regress out cell cycle genes to maintain biological variation. To estimate the activity of three gene groups as listed on Table S2, we applied AUCell method (Aibar et al., 2017). The regulation activity, ranging from 0 to 1, is the enrichment of the genes in the input gene set across the expression ranking of all genes in a cell. The three sets include "core pluripotency markers", "Naïve pluripotency markers", and "Other pluripotency markers". A cell with a high gene set score suggests that the gene set actively regulate transcriptional state in the cell. As shown in Figure S5, no difference was observed in the activity of the three gene sets in the four samples, suggesting that the StemFlex and E8 equally maintain pluripotency states in the cell.

Statistical analysis. Data are expressed as mean ± standard error of the mean (SEM). All statistical analyses and graphical data were generated using Graphpad Prism software (v6, www.graphpad.com). Statistical methods utilised were two-tailed t-test and two-way ANOVA test followed by Sidak's multiple comparisons test. Statistical significance was established from $p<0.05$.

Supplemental References

Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.-C., Geurts, P., Aerts, J., van den Oord, J., Atak, Z.K., Wouters, J., Aerts, S., 2017. SCENIC: single-cell regulatory network inference and clustering. Nat. Methods 14, 1083–1086.

Anders, S., Huber, W., 2010. Differential expression analysis for sequence count data. Genome Biol. 11, R106.

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4, 7.

Colella, S., Yau, C., Taylor, J.M., Mirza, G., Butler, H., Clouston, P., Bassett, A.S., Seller, A., Holmes, C.C., Ragoussis, J., 2007. QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. Nucleic Acids Res. 35, 2013–2025.

Crombie, D.E., Daniszewski, M., Liang, H.H., Kulkarni, T., Li, F., Lidgerwood, G.E., Conquest, A., Hernández, D., Hung, S.S., Gill, K.P., De Smit, E., Kearns, L.S., Clarke, L., Sluch, V.M., Chamling, X., Zack, D.J., Wong, R.C.B., Hewitt, A.W., Pébay, A., 2017. Development of a Modular Automated System for Maintenance and Differentiation of Adherent Human Pluripotent Stem Cells. SLAS Discov 2472555217696797.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21.

Fuchsberger, C Abecasis, GR Hinds, DA, n.d. minimac2: faster genotype imputation. Bioinformatics 31, 782–784.

Garrison, E., Marth, G., 2012. Haplotype-based variant detection from short-read sequencing. arXiv 1207.3907v2 [q–bio.GN].

Kang, H.M., Subramaniam, M., Targ, S., Nguyen, M., Maliskova, L., Wan, E., Wong, S., Byrnes, L., Lanata, C., Gate, R., Mostafavi, S., Marson, A., Zaitlen, N., Criswell, L.A., Ye, J., 2017. Multiplexing droplet-based single cell RNA-sequencing using natural genetic barcodes. https://doi.org/10.1101/118778

Leys, C., Ley, C., Klein, O., Bernard, P., Licata, L., 2013. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. J. Exp. Soc. Psychol. 49, 764–766.

Li, H., 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27, 2987–2993.

Loh, P.-R., Danecek, P., Palamara, P.F., Fuchsberger, C., A Reshef, Y., K Finucane, H., Schoenherr, S., Forer, L., McCarthy, S., Abecasis, G.R., Durbin, R., L Price, A., 2016. Reference-based phasing using the Haplotype Reference Consortium panel. Nat. Genet. 48, 1443–1448.

Lun, A.T.L., Bach, K., Marioni, J.C., 2016. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. Genome Biol. 17, 75.

McCaughey, T., Liang, H.H., Chen, C., Fenwick, E., Rees, G., Wong, R.C.B., Vickers, J.C., Summers, M.J., MacGregor, C., Craig, J.E., Munsie, M., Pébay, A., Hewitt, A.W., 2016. An Interactive Multimedia Approach to Improving Informed Consent for Induced Pluripotent Stem Cell Research. Cell Stem Cell 18, 307–308.

O'Brien, C.M., Chy, H.S., Zhou, Q., Blumenfeld, S., Lambshead, J.W., Liu, X., Kie, J., Capaldo, B.D., Chung, T.-L., Adams, T.E., Phan, T., Bentley, J.D., McKinstry, W.J., Oliva, K.,

McMurrick, P.J., Wang, Y.-C., Rossello, F.J., Lindeman, G.J., Chen, D., Jarde, T., Clark, A.T., Abud, H.E., Visvader, J.E., Nefzger, C.M., Polo, J.M., Loring, J.F., Laslett, A.L., 2017. New Monoclonal Antibodies to Defined Cell Surface Proteins on Human Pluripotent Stem Cells. Stem Cells 35, 626–640.

Okita, K., Matsumura, Y., Sato, Y., Okada, A., Morizane, A., Okamoto, S., Hong, H., Nakagawa, M., Tanabe, K., Tezuka, K.-I., Shibata, T., Kunisada, T., Takahashi, M., Takahashi, J., Saji, H., Yamanaka, S., 2011. A more efficient method to generate integration-free human iPS cells. Nat. Methods 8, 409–412.

Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S.F.A., Hakonarson, H., Bucan, M., 2007. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res. 17, 1665–1674.

Wu, G., Dawson, E., Duong, A., Haw, R., Stein, L., 2014. ReactomeFIViz: the Reactome FI Cytoscape app for pathway and network-based data analysis. F1000Res. https://doi.org/10.12688/f1000research.4431.1

Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., Gregory, M.T., Shuga, J., Montesclaros, L., Underwood, J.G., Masquelier, D.A., Nishimura, S.Y., Schnall-Levin, M., Wyatt, P.W., Hindson, C.M., Bharadwaj, R., Wong, A., Ness, K.D., Beppu, L.W., Deeg, H.J., McFarland, C., Loeb, K.R., Valente, W.J., Ericson, N.G., Stevens, E.A., Radich, J.P., Mikkelsen, T.S., Hindson, B.J., Bielas, J.H., 2017. Massively parallel digital transcriptional profiling of single cells. Nat. Commun. 8, 14049.